Biophysical and biochemical characterisation of DNA-based inhibitors of the cytosine-mutating APOBEC3 enzymes

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Biochemistry

Massey University, Palmerston North,

New Zealand

Fareeda Maged Yahya Mohammad Barzak

2020

Abstract

With the rise of antiviral and anticancer drug resistance, a new approach must be taken to overcome this burden. The APOBEC3 (A3) family of cytosine deaminases hypermutate cytosines to uracils in single-stranded DNA (ssDNA). These enzymes act as double-edged swords: on one side they protect humans against a range of retroviruses and other pathogens, but several A3s are exploited by viruses and cancer cells to increase their rate of evolution using the enzyme's mutagenic actions. This latter mode permits escape of cancer cells from the adaptive immune response and leads to the development of drug resistance. In particular, APOBEC3B (A3B) is considered to be a main driving source of genomic mutations in cancer cells. Inhibition of A3B, while retaining the beneficial actions of the other A3 in the immune system, may be used to augment existing anticancer therapies.

In this study, we showed for the first time that short ssDNAs containing cytosine analogue nucleosides, 2'-deoxyzebularine (dZ) or 5-fluoro-2'-deoxyzebularine (5FdZ) in place of the substrate 2'-deoxycytidine (dC) in the preferred 5'-TC motif, inhibit the catalytic activity of A3B. However, as most A3 enzymes (except A3G) prefer to deaminate ssDNA with a 5'-TC motif, selective A3B inhibition was uncertain. We noted that nucleotides adjacent to the 5'-CCC motif influence the dC deamination preference of A3A, A3B, and A3G's. Replacement of the A3B's preferred dC in the 5'-CCC motif with dZ (5'-dZCC) led to the first selective inhibitor of A3B, in preference to A3A and A3G. Furthermore, using small-angle X-ray scattering (SAXS) we obtained the first model of a full-length two-domain A3 in complex with a dZ-ssDNA inhibitor. Our model showed that the ssDNA was largely bound to the C-terminal domain (CTD) with limited contact to the N-terminal domain in solution, due to the high affinity of the dZ for the CTD active-site. Our work provides a new platform for use of ssDNA-based inhibitors in targeting the mutagenic action of the A3B. Further developments using more potent inhibitors will help to achieve inhibition in cellular studies, with the ultimate goal to complement anti-cancer and antiviral treatments.

Acknowledgements

"If I have seen further, it is by standing on the shoulders of giants." – Isaac Newton. Therefore, I would like to acknowledge the people who made this project possible.

First and foremost, I wish to express my sincere gratitude to my advisors, Assoc. Prof. Vyacheslav Filichev, Dr. Elena Harjes, and Prof. Geoffrey Jameson for giving me the opportunity to undertake this research and expertly guiding me throughout this project. Their continuous support, patience, advice, data analysis, and insightful discussions made this project possible.

I would like to extend a big thanks to our APOBEC team for the endless discussions, ideas, and technical support. Dr. Maksim Kvach and Mr. Harikrishnan Mohana Kurup for their design and synthesis of our chemically modified-oligonucleotides, and Dr. Stefan Harjes for his supporting work on A3A and A3G. In addition, thanks to our X-lab structural biology group for their stimulating discussions and advice.

Thank you to all the SFS staff who have facilitated my studies through the years, especially Prof. Kathryn Stowell and Assoc. Prof Gillian Norris. A special thanks goes to the lab ladies' group; your friendships and support have meant the world to me.

I am deeply grateful for the assistance of Dr. Timothy Ryan from the Australian Synchrotron SAXS/WAXS beamline for help with SAXS analysis and deconvolution of A3G data. Thank you to Dr. Patrick Edwards for assistance with the NMR spectrometer and Mr. Trevor Loo for performing mass spectrometry experiments and all technical assistance.

I wish to acknowledge all our collaborators. Prof. Reuben Harris and Assoc. Prof. Hideki Aihara along with their teams from the University of Minnesota for kindly providing all the plasmid constructs used in this study and for technical training during my visit, Dr. Linda Chelico and her team from the University of Saskatchewan for kindly providing full-length A3G protein used in SAXS studies, Assoc. Prof. Daniel Harki and his team from the University of Minnesota for supporting inhibitor studies on the human A3A and $A3B_{CTD}$ enzymes, and Prof. Kurt Krause and his team from the University of Otago for providing full-length A3G for inhibition assays used to design the A3G oligonucleotides for SAXS studies.

Thank you to the School of Fundamental Sciences (SFS) for funding my research under the graduate assistance scholarship and their travel fund to present my research at ComBio2018 conference in Sydney. Thank you to NZMS for student travel grant aid to present my research at the NZMS/NZSBMB microbes and molecules meeting 2018 in Dunedin. Furthermore, thank you to both the student travel grants from the Australian Synchrotron and NZSBMB for allowing me to attend and present my work at the Australian Synchrotron User Meeting 2018 in Melbourne.

Thank you to my friends, siblings, and family for their support. Most importantly, I wish to thank my parents Maged and Sohair for their encouragement, love, support, and patience throughout my life. I am forever indebted for everything you provided for me. I could not have done any of this without you. I dedicate this thesis to you.

List of Abbreviations

General abbreviations and definitions

Abbreviation	Definition
°C	Degrees Celsius
× g	Multiples of gravitational force
Δ*G*	Gibbs free energy
Δ*H*	Enthalpy
Δ*S*	Entropy
ΔT_m	Change in melting temperature
λ	Wavelength
^{1}H-NMR	Proton NMR
$1/V_0$	Inverse of initial rate of a reaction
1/[S]	Inverse of the substrate concentration
1/[I]	Inverse of the inhibitor concentration
298 K	298 Kelvin
5FdZ	5-Fluoro-2'-deoxyzebularine
Å	Angstrom (10^{-10} m)
A_{260}	Absorbance at 260 nm
A_{280}	Absorbance at 280 nm
A_{600}	Absorbance at 600 nm
A1	APOBEC1
A2	APOBEC2
A3(A-H)	APOBEC3A-H
A3A-mimic	$A3B_{CTD}$-QM-ΔL3-AL1swap
$A3B_{CTD}$-AL1	$A3B_{CTD}$-QM-ΔL3-AL1swap
$A3B_{CTD}$*	A3BCTD-QM-ΔL3-A3Aloop1swap-E255A
A4	APOBEC4
aa	Amino acid
ABC	Ammonium bicarbonate
ACN	Acetonitrile

Abbreviation	Definition
AID	Activation-induced cytosine deaminase
Amp	Ampicillin
APOBEC	Apolipoprotein B mRNA-editing enzyme-catalytic polypeptide
aq.	Aqueous
Au	Absorbance units
bp	Base pair
β-ME	β-mercaptoethanol
cDNA	Complementary DNA
CDA	Cytosine deaminase
Chi^2 (or χ^2)	Chi-squared distribution
CTD	Carboxyl terminal domain
C-terminal	Carboxyl terminal
D_2O	Deuterium oxide
dA	2'-Deoxyadenosine
Da	Daltons
dadU	2'-Deoxy-3-deazauridine
dadZ	2'-Deoxy-3-deazazebularine
dC	2'-Deoxycytidine
dG	2'-Deoxyguanosine
D_{max}	Maximum dimension
dT	2'-Deoxythymidine
dU	2'-Deoxyuridine
dZ	2'-Deoxyzebularine
DNA	2'-Deoxyribonucleic acid
dsDNA	Double-stranded DNA
DSS	4,4-Dimethyl-4-silapentane-1-sulfonic acid
DTT	Dithiothreitol
[E]	Enzyme concentration
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EFA	Estimated factor analysis

Abbreviation	Definition
His_6	Hexa-histidine tag
HMW	High molecular weight
HSQC	Heteronuclear single quantum coherence
I	Scattering intensity
[I]	Inhibitor concentration
$I(0)$	Forward scattering intensity
Ig	Immunoglobulin
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl-β-D-thiogalactoside
ITC	Isothermal titration calorimetry
K_a	Equilibrium binding association constant
kb	Kilobase pairs (of DNA)
k_{cat}	Catalytic rate constant
k_{cat}/K_m	Specificity constant (catalytic efficiency)
K_d	Equilibrium dissociation binding constant
kDa	KiloDalton
K_i	Inhibition constant
K_m	Michaelis constant (apparent binding)
LB	Luria Bertani media
LMW	Low molecular weight
M9 media	M9 minimal media
mAU	Milli absorbance units
min	Minute
mM	Millimolar
mRNA	Messenger RNA
MW	Molecular weight
n	Sample size
NES	Nuclear export signal
NLS	Nuclear localisation signal
NMR	Nuclear magnetic resonance
NSD	Normalised spatial discrepancy

Abbreviation	Definition
N-terminal	Amino terminal
NTD	N-terminal domain
OD	Optical density at 600 nm
Oligo	Oligonucleotide
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein data bank
pH	Negative decadal logarithm of proton activity (concentration)
PIPES	1,4-Piperazinediethanesulfonic acid
$P(r)$	Pair distribution plot
p.s.i.	Pounds per square inch
q	Scattering vector
q_{min}	Minimum scattering vector
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
Rg	Radius of gyration
RT	Reverse transcriptase
[S]	Substrate concentration
SAXS	Small angle X-ray scattering
sec	Seconds
SEC	Size exclusion chromatography
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
ssDNA	Single-stranded DNA
SVD	Singular value decomposition
TB	Transformation buffer
TEMED	Tetramethylethylenediamine
T_m	Melting temperature
Tris	Trisaminomethane
TWEEN 20	Polyethylene glycol sorbitan monolaurate

Abbreviation	Definition
V	Porod volume
V_0	Initial rate of a reaction
V_{max}	Maximum rate of a reaction
Wt	Wild-type

Oligonucleotides used in this study

Abbreviations	Sequence
9merC-oligo or TCA-oligo	5'-ATTTCATTT
9merU-oligo	5'-ATTTUATTT
9ATC-oligo	5'-AAATCAAAA
9ATU-oligo	5'-AAATUAAAA
A9C-oligo	5'-AATTCAAAA
A9U-oligo	5'-AATTUAAAA
CCC-oligo	5'-ATTCCCAATT
CCU-oligo	5'-ATTCCUAATT
CUC-oligo	5'-ATTCUCAATT
UCC-oligo	5'-ATTUCCAATT
CUU-oligo	5'-ATTCUUAATT
UUU-oligo	5'-ATTUUUAATT
TC-3' 6-FAM-oligo	5'-ATTATTATTATTCAAATGGAT TTATTTATTTATTTATTTATTT-fluorescein
dZ-oligo	5'-ATTT**dZ**ATTT
5FdZ-oligo	5'-ATTT**5FdZ**ATTT
3dadZ-oligo	5'-ATTT**3dadZ**ATTT
3dadU-oligo	5'-ATTT**3dadU**ATTT
dZCC-oligo	5'-ATT**dZ**CCAATT
CCdZ-oligo	5'-ATTCC**dZ**AATT
$CCdZT_{30}$-oligo (termed CCdZ-40mer)	5'-ATTCC**dZ**AATTT$_{30}$

Amino acid names and abbreviation

Amino acid	3-letter abbreviation	1-letter abbreviation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Table of contents

List of Figures

List of Tables

1. Introduction

This chapter provides a brief background on the human APOBEC family, focusing on the APOBEC3 subfamily structure, functions, and pathological consequences. I also delineate the aims of this study and rationale of investigation.

1.1 Preface

Over the last millennium, exponential advancements have been made in the treatment of various cancers. From ancient Egyptian civilisation until the 18th century, surface tumour masses were removed by radical surgical and burning procedures, but this led to poor prognosis and life expectancy due to poor hygiene practices [30, 254]. By the 19th century, surgical practices improved and led to better survival rates [72]. The discovery of X-rays in the late 19th century lead to the development of novel non-surgical radiation therapy to treat tumours [66, 173]. A major breakthrough came in 1956 when the first chemotherapy drug was discovered capable of targeting a large array of cancers [44]; then by the 1980s specific anticancer drugs targeting particular tumours were developed [130]. These therapeutic treatments provided improved patient outcomes but also led to toxicity of normal cells. However, it became apparent that tumours could develop resistance to these standard therapies [44]. Tumour cells can undergo changes through a number of clonal expansion, genetic diversification, and selective pressures leading to the evolution of cancer and evasion of therapy [85-86, 288].

Over the last two decades, immunotherapeutic and monoclonal antibody drugs have been utilised in conjunction with standard therapies against the advancement of tumours and metastasis, obtaining better treatment outcomes [229, 263]. Eliminating the driving factors in cancer evolution would be essential to prevent the development of drug resistance and in turn may enhance the efficacy of standard therapies. Recent studies, indicated that a major driving force behind somatic mutations in various cancers was by the action of the APOBEC3B (A3B) enzyme, found in the nucleus [12, 34, 137-138, 147, 253]. The A3B belongs to a family of cytosine deaminases (APOBEC3), which mutate cytosines to uracils in single-stranded DNA (ssDNA). In normal cells, the A3 enzymes defend the humans against a range of retroviruses and pathogenic viruses. However, in cancer cells A3B is upregulated, enabling the accumulation of somatic mutations and progression of tumours in line with resistance against anticancer drugs. Selective inhibition of A3B, which has not been investigated previously, would provide a method in preventing cancer evolution and the development of drug resistances, while maintaining the beneficial actions of the other A3 enzymes.

Substrate 2'-deoxycytidine (dC) within specific deamination sequences in ssDNA binds into the active site of the A3 enzyme, where it is deaminated to 2'-deoxyuridine (dU). As the A3 enzymes prefer to interact with ssDNA [15, 26, 48], we aimed to design and evaluate substrate-like ssDNAs incorporated with cytosine analogue nucleosides in place of dC in the preferred A3B deamination sequence motifs (5'-TC, and 5'-CCC, underlined is target dC). By placing these cytosine analogue nucleosides in these ssDNA sequences, the ssDNA would rationally deliver these nucleosides into the active site of the enzyme. These cytosine analogues are anticipated to mimic the transition state of deamination and would remain bound in the active site, thereby inhibiting the mutagenic action of A3B. The development of these potential inhibitors offers a promising new approach to complement existing anticancer therapies.

1.2 Zinc-dependent cytosine deaminases

APOBEC enzymes are zinc-dependent cytosine deaminases that catalyse the conversion of cytosines to uracils (C-to-U) through a hydrolytic reaction as shown in Figure 1.1 [100, 113, 210]. A closely related ancestral enzyme cytosine deaminase (CDA) functions through the pyrimidine salvage pathway by deaminating free cytosines to recover uracils [296].

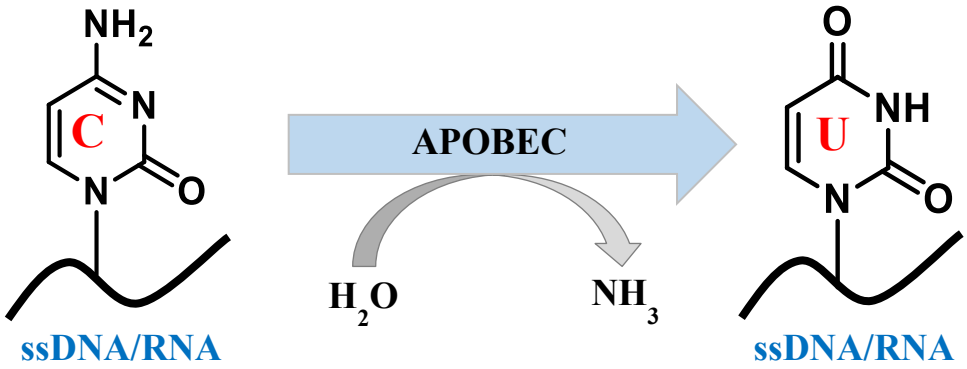


Figure 1.1 APOBEC catalysed cytosine deamination of polynucleotides.
Cited with permission Springer Nature [210], copyright (2013).

Like the ancestral CDA, all the APOBEC members contain a conserved cytosine deaminase motif within the active site that facilitates the deamination of the target cytosines [63, 100, 142, 276] (discussed further 1.6.2). The evolutionary conservation of this motif implies that the general method of cytosine deamination by APOBEC enzymes is tentatively analogous to that of the reported mechanism of CDA [14, 120]. However, due to structural differences the APOBEC family members tend to deaminate cytosines within polynucleotides such as single-stranded DNA (ssDNA) or RNA over single cytosine

nucleosides or nucleotides, as further discussed in section 1.6.3.1 and 1.6.3.2) (Figure 1.1).

1.3 The human APOBEC family

In humans, the APOBEC family consists of eleven distinct members that are grouped together based on sequence homology. Members of this family include the activation induced deaminase (AID), APOBEC1 (A1), APOBEC2 (A2), APOBEC4 (A4), and seven APOBEC3 (A3A-H) enzymes, which are encoded as single- or double-domain entities on various chromosomes [100, 113] (Figure 1.2).

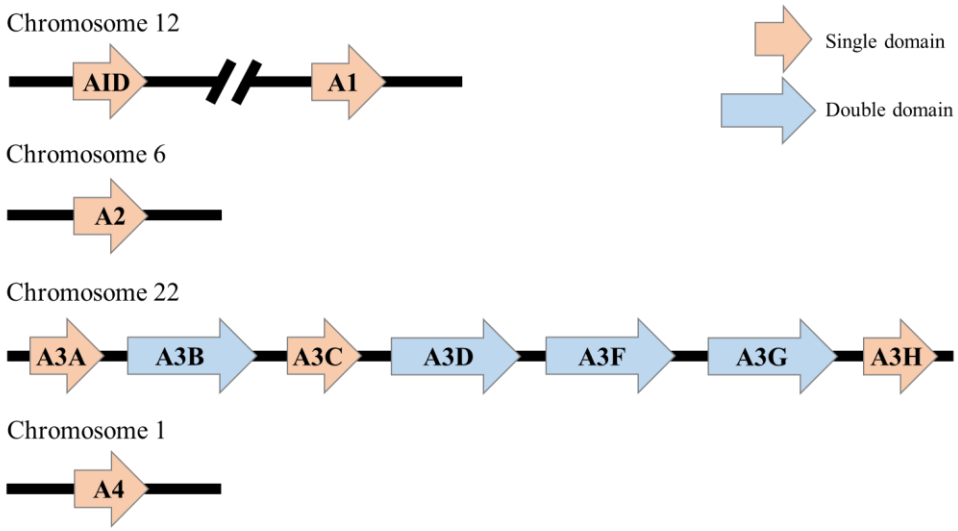


Figure 1.2 APOBEC family members chromosomal arrangement.
Cited with permission Springer Nature [100], copyright (2004).

The members derive their name from the founding member, the apolipoprotein B (apoB) mRNA-editing catalytic enzyme 1 (A1), which is the only enzyme that has the ability to deaminate cytosines within the apolipoprotein B (apoB) pre-mRNA for the regulation of lipid metabolism [60, 95, 142, 186, 262]. Other APOBEC members, such as AID and A3 subfamily, have been characterised to act in various roles within the immune system, largely by deaminating ssDNA [41, 95, 122-123, 207, 210, 252]. The catalytic action of these enzymes must be tightly regulated to prevent detrimental effects on the host.

1.3.1 APOBEC1 (A1)

The first APOBEC family member identified, A1, is mainly responsible for regulating the secretion of lipoproteins by editing the apoB pre-mRNA [186, 262]. Recent *in vitro* studies have implied that A1 can also target various RNAs [19, 215], while its overexpression encompasses DNA deaminase activity with no known substrates, but maybe related to

the other APOBEC members [17, 95, 106, 202-203]. A1 is encoded by a gene on chromosome 12, separated by nearly 1 megabase from the *AID* gene (Figure 1.2) and is expressed in the gastrointestinal tissues and testis [182].

The A1 enzyme specifically deaminates cytosine 6666 (C6666) of the apolipoprotein B (apoB100) pre-mRNA, which changes a glutamine codon into a stop codon, leading to a truncated form of apoB (apoB48). The generated apoB48 form is the main component of the lipoprotein chylomicron, which is necessary for lipid transport [2, 204, 262]. This enzyme was described to have an intrinsic dinucleotide binding preference of 5'-TC [113], but it also binds to RNA with low affinity [2, 184-185]. Therefore, RNA binding proteins, APOBEC1 complementation factor (A1CF) and an RNA-binding motif protein 47 (RBM47), binds upstream of the target C6666 to recruit A1 to the target mRNA achieving recruitment of A1 and providing specific deamination [73, 81]

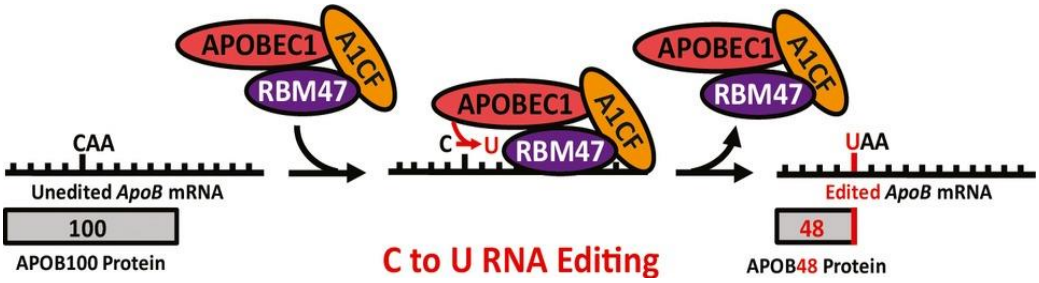


Figure 1.3 A1 deamination of apoB100 pre-mRNA aided by the RNA binding proteins A1CF and RBM47.

Reprinted with permission from John Wiley and Sons [73], copyright (2014).

It was noted that the overexpression of A1 *in vitro* was linked to oncogenesis [285]. Therefore, to prevent this potential A1's catalytic activity appears to be tightly controlled through a number of mechanisms. First, the presence of the RNA binding proteins, described above, regulate the binding of A1 to the target substrate (Figure 1.3). Second, the A1 enzyme functions in the nucleus, but it is shuttled out to the cytoplasm by a carboxy-terminal nuclear export signal (NES) when not required and back into the nucleus by an amino-terminal nuclear localisation signal (NLS) when required [144, 287]. Finally, the expression of A1 is modulated to specific tissues [249].

1.3.2 Activation-induced cytosine deaminase (AID)

AID is also encoded by a gene on chromosome 12 and is capable of nucleotide editing (Figure 1.2) [113]. Initially, it was thought that AID was an RNA-editing enzyme as it has the closest sequence homology to A1. However, unlike A1 it was discovered to largely act as a DNA mutator [201]. This enzyme is specifically expressed in germinal centres of activated B cells and mainly functions in the adaptive immune system through antibody diversification against a range of pathogens [28, 176-177]. Antibody maturation is accomplished when AID deaminates the complementary DNA (cDNA) of the immunoglobulin (Ig) locus variable regions leading to the accumulation of mutations (changes C:G pairs into U:G mispairs). As shown in Figure 1.4, this initiates the base-repair mechanism where uracil DNA glycosylase (UDG) cleaves uracil leaving abasic sites, triggering somatic hypermutation, class-switch recombination, or gene conversion, which modifies the antibodies' affinity and isotypes [4, 95, 104, 175-176, 188].

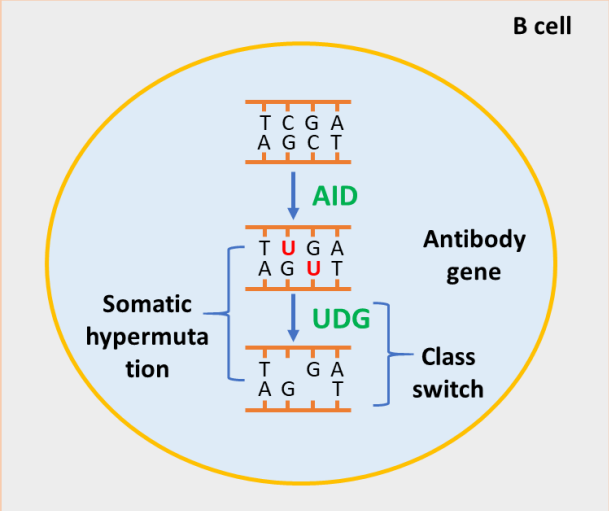


Figure 1.4 AID deaminase and antibody diversification.

Cited with permission from Springer Nature [210], copyright (2013).

Regulation is crucial to avoid genomic mutations arising due to the enzyme having DNA editing activity [192]. AID is acutely expressed in the activated B cell tissues, but, similar to A1, its deaminase activity is controlled through multiple mechanisms. First, AID is preferentially cytoplasmic and localises to the nucleus only when required [28, 109]. Second, the nuclear lifespan of AID is dictated by ubiquitination-dependent degradation [3]. Finally, AID is thought to selectively target its preferred dinucleotide in ssDNA when a cytosine is preceded by a purine (A/G) [41, 123] by binding to cis-acting element in the Ig loci or other cofactors [199, 260]. Lately, it has also been reported that AID is involved in the innate

immune system. AID binds directly to the mRNA of some retroelements, to prevent these elements from inserting their DNA into the host genome, but the mechanism remains unclear [160, 169].

1.3.3 APOBEC2 (A2)

A less known APOBEC enzyme, A2 is encoded by a gene on chromosome 6 [100] (Figure 1.2). A2 was observed to be expressed in cardiac and skeletal muscles [153], noted to be essential for the maintenance and development of muscles [222, 271]. Earlier studies implied that this enzyme was unable to target single-stranded polynucleotides [95, 153]. Through its close sequence homology and similar crystal structure to APOBEC members it can be assumed that A2 is capable of binding polynucleotides [134, 205]. Currently, an experimental study has depicted that A2 targets specific tumour suppressor genes by editing RNA [193]. However, the mechanism and significance of A2 remain to be unearthed.

1.3.4 APOBEC4 (A4)

A4 was found to have the lowest sequence homology to the APOBEC family members [62, 214]. It is encoded by a gene on chromosome 1 and expressed in the human testis, suggesting that it may be involved in spermatogenesis [163]. Lately, *in vitro* experiments indicated that A4 may enhance the expression of HIV-1 by mediating the activity of the promotor [163]. Several studies found that A4 lacked mutagenic activity [171, 210] and the APOBEC signature motif portrayed in section 1.6.2, suggesting that it may function through an alternative mechanism that is yet to be discovered.

1.3.5 APOBEC3 subfamily

Unlike other species, humans contain seven A3 enzymes (A3A-A3H) that share similar sequence homologies and are encoded by genes clustered in tandem on chromosome 22 [100, 113, 143]. These enzymes play significant roles in the innate immune system, and it is believed that in humans they diverged and expanded from a single enzyme to permit the protection against a range of exogenous viruses and endogenous retroelements, as shown in Figure 1.5 and further discussed in section 1.4 [97-98, 100, 113, 210].

The A3 enzymes are regulated through a number of mechanisms (refer to section 1.5) and accomplish their functions largely by modifying viral ssDNA [49, 95, 113, 230] (discussed in section 1.6.3.2). A3A and A3G have recently been identified to be capable of editing RNA [234-236], but with less efficiency than ssDNA targets (unpublished raw data, Dr. Stefan Harjes). However, misregulation of A3 activity is linked to pathological consequences, such as viral and cancer evolution further discussed in section 1.7 [34-35, 138, 140, 255] (Figure 1.5).

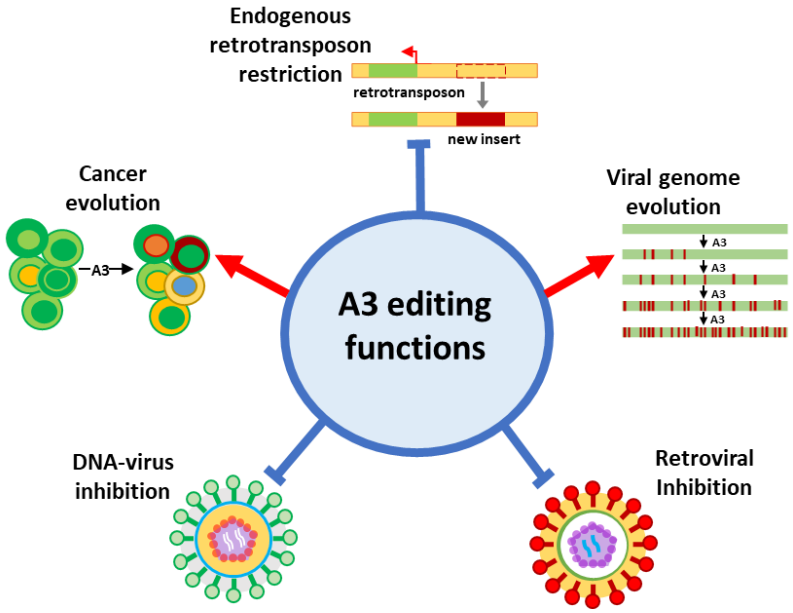


Figure 1.5 Roles of A3 deamination activity.

Cited with permission from Elsevier [119], copyright (2016).

1.4 The A3 immune system defence mechanisms

1.4.1 Restriction of retroviruses

Typically, upon infection of a host cell with the retrovirus, the hosts' reverse transcriptase (RT) generates a cDNA intermediate from the retroviral RNA in the cytoplasm. The cDNA is then reverse transcribed using the same enzyme eventually forming a double-stranded DNA (dsDNA) which then migrates into the nucleus as displayed in Figure 1.6. In the nucleus the dsDNA is integrated into the host cell's DNA where it can replicate using the host cell's machinery and the cycle of infection can repeat. Several A3 enzymes have been identified to restrict numerous retroviruses, such as simian immunodeficiency viruses (SIVs), murine leukaemia virus (MLV), human T-cell lymphotropic virus (HTLV), by deaminating cytosines within the nascent retroviral cDNA intermediate. The pathogenic lentivirus human immunodeficiency virus type-1 (HIV-1) has been

extensively studied and was one of the first discovered retroviruses to be inhibited by A3G [98, 148, 237]. In addition to A3G, three other APOBECS, A3D, A3F, and A3H, were also identified to be capable of restricting the replication of HIV-1 [211, 274]. Several assays noted that A3G and A3F were co-expressed and were the most potent inhibitors of HIV-1, suggesting their mutual role, while A3D and A3H inhibited HIV-1 modestly [17, 141, 278].

In infected T lymphocytes lacking the counter restrictive virion infectivity factor (Vif) (discussed further in next paragraphs), the A3G/A3D/A3F/A3H present in the cytoplasm interact with the virus group-specific antigen (Gag) ribonucleoprotein complex allowing for its encapsulation along with the HIV genomic RNA into a budding virion [20, 88, 110, 165, 258]. Once the virion attaches to its target cell, its components are released and the viral genomic RNA undergoes reverse transcription producing the nascent viral cDNA, which is quickly deaminated by the A3 enzyme resulting in C-to-U hypermutations [98, 290-291]. The mutated cDNA is then either targeted for degradation or is used as a template for the synthesis of the second DNA strand, which compromises the virus' viability to subsequently replicate due to potentially lethal non-functional mutations (second strand contains G-to-A hypermutations). as shown in Figure 1.6 [98, 148, 237]. Alternatively, HIV-1 replication can also be prevented through a deaminase-independent method, where the A3s bind to the viral RNA and block the path of reverse transcriptase (RT) [16, 18, 110-111, 189].

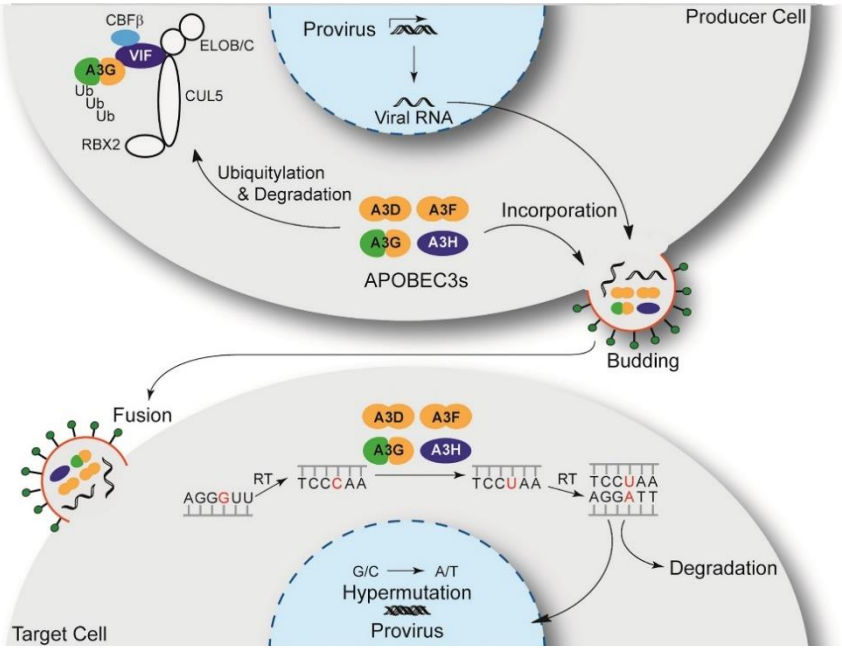


Figure 1.6 A3 inhibition of HIV-1 replication mechanism.

Reprinted with permission from Elsevier [195], copyright (2018).

However, retroviruses have developed several mechanisms to overcome this immune response or indeed to utilise the A3 action for viral diversification, which is further elaborated in section 1.7.1. One well-known mechanism to override the restriction of the HIV-1 retrovirus by the A3 enzymes is by the presence of a small protein termed the virion infectivity factor (Vif). Vif mediates ubiquitination by the E3 ligase and subsequent proteasomal degradation of the A3 enzymes, allowing the HIV-1 virus to infect and propagate in susceptible cells [61, 100, 162, 237] (Figure 1.6). Furthermore, Vif was noted to also be capable of stopping the action of A3s by preventing its encapsulation into the HIV virion and blocking its catalytic activity by means that are not well understood [42, 87, 238].

1.4.2 Targeting endogenous retrotransposons

Endogenous retrotransposons are mobile DNA elements that make up a large fraction of the human genome [139]. As displayed in Figure 1.7, these retroelements are first transcribed in the nucleus forming an RNA intermediate that is transferred into the cytoplasm. In the cytoplasm, proteins, such as ORF1 and ORF2, attach to the retroelement RNA forming a ribonucleoprotein (RNP) complex. This complex is translocated into the nucleus where it is reverse transcribed forming cDNA, which is inserted into a new site in the genome [198].

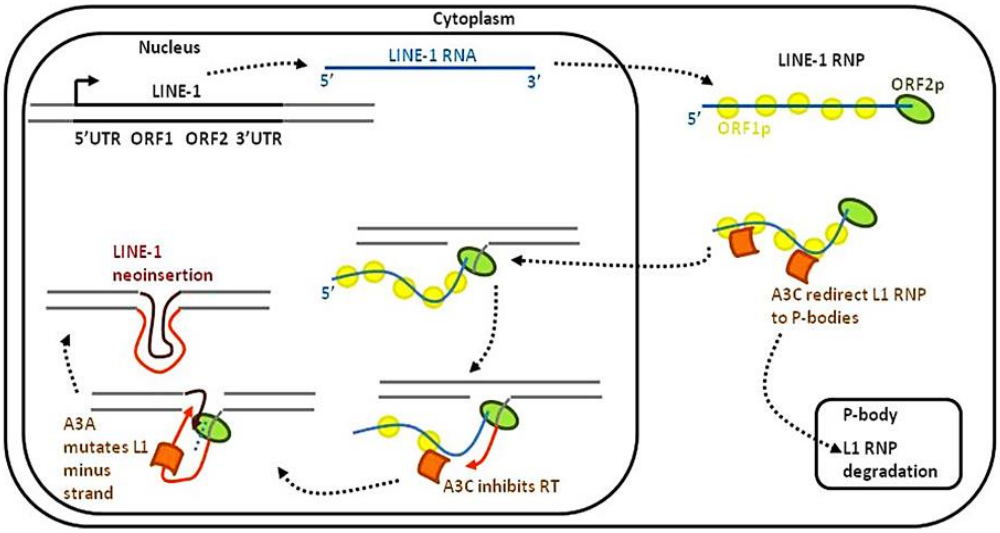


Figure 1.7 Replication of the retrotransposon LINE-1 and its inhibition by A3 enzymes.
Reprinted from open access article MDPI, Willems and Gillet [279], copyright (2015).

Many studies have indicated that these transposable elements play a key role in the evolution of the genome, genetic diversity, and regulation of gene expression. However, the ability of retrotransposons to reorganise genetic elements also comes at a cost, causing genetic instabilities that lead to detrimental effects such as various genetic disorders and

cancers [52, 64, 126, 139, 170]. Overtime, mechanisms have arisen to maintain the integrity of the genome by shaping the action of these elements. The expansion of the human A3 family has not only been prompted to target exogenous viruses, but it is now thought to have arisen to restrict or diversify the retroelements to benefit the host [221, 223, 292]. Several A3 members, including A3A, A3B, A3C, A3F, and A3G, have been implicated in the inhibition of endogenous retrotransposons, such as long terminal repeats (LTR), and non-LTR (such as long interspersed nuclear elements, LINE, and short interspersed nuclear elements, SINE) [54, 69-71, 141, 210]. A3 enzymes deaminate the reverse transcribed cDNA causing significant mutations in the retrotransposon. This can either inactivate further cycles of retrotransposition or decrease the number of transposed cDNA copies [71]. Inhibition of the retrotransposons can also be achieved through the mutation of the minus DNA strand by A3 or through deamination-independent mechanisms by directly binding to the intermediate RNA transcript and redirecting the complex to p-bodies for degradation (see Figure 1.7). However, the overall system remains unclear [54, 124].

1.4.3 Inhibition of DNA-based viruses

Aside from retroviruses, the A3 family has been implicated to restrict a large subset of DNA-based viruses. The cancer-associated viruses, hepatitis B virus (HBV), which replicates in the liver and human papillomavirus (HPV), which replicates in keratinocytes of the skin, are key targets of several A3 members. Unlike retroviruses (refer to 1.4.1), the viral DNA parasite is first transcribed into an intermediate pre-genomic RNA in the nucleus by the hosts machinery. In the cytoplasm, the pre-genomic RNA is reverse transcribed to an intermediate cDNA (minus strand) then to partially dsDNA, as shown in Figure 1.8.

In vitro assays noted that both DNA strands of HBV and HPV are deaminated by A3 enzymes during transcription or replication of dsDNA resulting in either C-to-T hypermutations (positive strand) or G-to-A hypermutations (negative strand) which together compromise the virus replication cycle [10, 121, 257, 266-267] (Figure 1.8). Furthermore, studies showed that the A3s can inhibit these viruses through deamination-independent

mechanisms by targeting the pre-genomic RNA, causing interference with reverse transcription [216, 265].

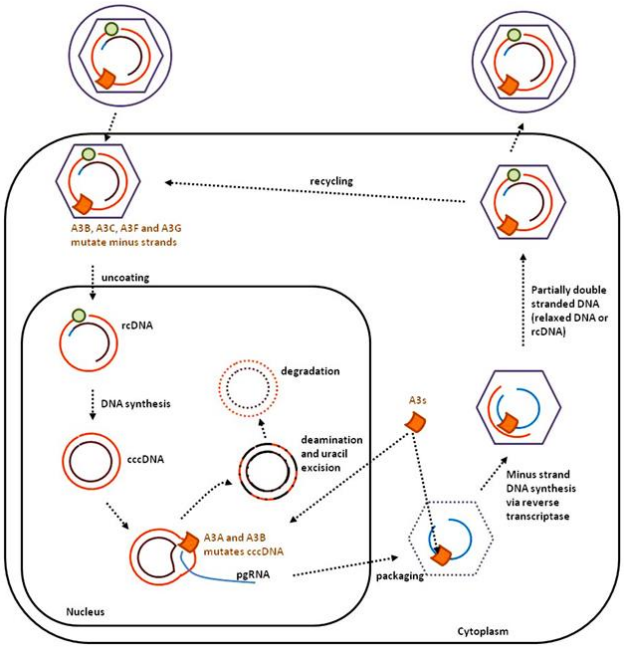


Figure 1.8 HBV DNA-based virus replication mechanism and inhibition by A3 enzymes.
Reprinted from open access article MDPI, Willems and Gillet [279], copyright (2015)

Several A3 members, A3B, A3C, A3G, and A3H, were found to cause low levels of HBV hypermutations in response to increased levels of cytokines released in the inflammatory response, leading to restriction of acute and chronic HBV infection [267]. Moreover, A3A, A3C, and A3H were shown *in vitro* to be involved in the hypermutations of the HPV genome [266, 275] to restrict HPV's wide range of diseases from warts to invasive cancers [29, 38]. However, studies noted a strong link between these DNA-based viruses and cancer progression, whereby the catalytic action of the A3s in cancer cells is hijacked [101, 191, 269, 275, 284] (described further in section 1.7.2).

1.5 Regulation of the A3 deamination activity

The capacity to control the A3s' functions relies on a number of mechanisms, including gene expression, post-transitional modifications, cofactors, tissue-specific expression, and subcellular distribution (Table 1.1) [125, 212]. Importantly, during mitosis all the A3 enzymes are excluded from the genomic DNA [137-138], but following mitosis A3B co-localises with the DNA in the nucleus, A3D, A3F, and A3G appear in the cytoplasm, while A3A, A3C, and A3H are widely distributed in both the cytoplasm and nucleus [12, 21, 137-138, 253] (Table 1.1). Subcellular regulation maybe a way to sequester the action of these enzymes away from the genomic DNA. However, A3 enzymes that are in contact

with the nucleus may have additional mechanisms to prevent a genotoxic effect on the cell.

Table 1.1 APOBEC family localisation and function.

Cited from [60, 100]

Name	Genomic location	Deaminase domains	Expression	Subcellular localisation	Editing activity	Function
AID	12	1	Activated B cells, testis	Cytoplasmic - acts in nucleus	DNA	Antibody diversification, retrotransposons
A1	12	1	Small intestine	Cytoplasmic - acts in nucleus	RNA, DNA	ApoB pre-mRNA
A2	6	1	Cardiac and skeletal muscles	Cytoplasmic /nuclear	Unknown	Unknown
A3A	22	1	Keratinocytes, blood	Cytoplasmic /nuclear	DNA (minor RNA)	DNA-viruses, retrotransposons
A3B	22	2	Peripheral-blood cells, T cells and keratinocytes	Nuclear	DNA	Retroviruses, DNA-viruses, retrotransposons
A3C	22	1	Multiple tissues	Cytoplasmic /nuclear	DNA	Retroviruses, DNA-viruses, retrotransposons
A3DE	22	2	Thyroid, spleen, blood	Cytoplasmic	DNA	Retroviruses
A3F	22	2	Multiple tissues	Cytoplasmic	DNA	Retroviruses, DNA-viruses, retrotransposons
A3G	22	2	Multiple tissues	Cytoplasmic	DNA (minor RNA)	Retroviruses, DNA-viruses, retrotransposons
A3H	22	2	Blood, thyroid, placenta	Unknown	DNA	DNA-viruses
A4	1	1	Testis	Unknown	Unknown	Unknown

Multiple studies have also implied the potential of oligomerisation for some A3 proteins (A3B, A3D, A3F-H), which may occur through protein self-association or an RNA-dependent manner [48, 65, 80, 103, 145, 150, 167, 205-206, 220, 250, 276, 293], although the necessity for multimerisation remains unresolved. Several studies have suggested that the action of multimerisation through surface interactions may regulate the A3s' catalytic activity by forming inactive conformations [22-23, 27, 79, 118, 233]. Therefore, to better understand the role of A3 oligomerisation solution-state structures would be useful.

1.6 Structural and biochemical insight of the A3 family

1.6.1 Organisation of the A3 domains

The A3 members are expressed as either a single or double domain. A3A, A3C, and A3H are characterised as single-domain enzymes, whereas A3B, A3G, A3D, and A3F are made up of two tandem domains believed to be linked by a short flexible linker [60, 142-143] (see linker in black connecting domains in Figure 1.9). In double domain A3 enzymes, the carboxy-terminal domain (referred to as the C-terminal domain or CTD) demonstrates catalytic activity, whereas the amino-terminal domain (referred to as the N-terminal domain or NTD) is said to be pseudo-catalytic but is believed to be important in binding substrates, which may regulate catalytic activity, consistent with observations that double-domain (referred to as full-length, fl) A3s are more active than the CTD alone [24, 78, 88, 91, 187, 189, 247].

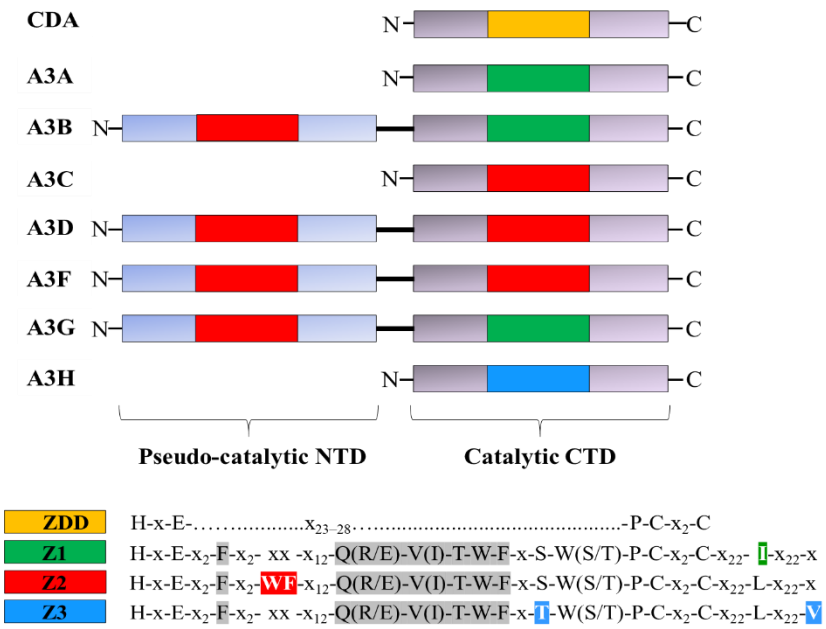


Figure 1.9 Organisation of the domains of the A3 family.

Cited with permission from Elsevier [220], copyright (2016).

In contrast to the zinc-dependent deaminase domain (ZDD) of the ancestral CDA and other cytosine deaminases, the domains of the A3 members can be further classified into subtypes (Z1, Z2, and Z3) based on specific amino acids found in the ZDD, highlighted (green, red, and blue) in Figure 1.9. These subclasses may explain the variation in substrate preference and function. The single-domain A3 enzymes are defined by a different Z-domain subtype (A3A-Z1, A3C-Z2, A3H-Z3), whereas the double-domain A3B and A3G are composed of Z2-Z1 domains, and the A3D and A3F consist of Z2-Z2 domains.

1.6.2 Structure of the A3 domain

Before 2015, there were no structures of the A3 enzymes so they were modelled using closely related zinc-dependent deaminases [14, 120, 158, 205]. Several single A3 domain structures have in the last few years been reported including A3A [36, 239], $A3B_{CTD}$ [37, 240], $A3B_{NTD}$ [282], A3C [118], $A3F_{CTD}$ [22, 232], $A3G_{CTD}$ [53, 79, 92, 103, 233], $A3G_{NTD}$ [128], and A3H [108, 231]. These all showed an overall similar domain structural fold despite differences in amino acid sequences. However, at the time this thesis research was conducted, structures of two-domain A3s were yet unknown.

Each domain shares a conserved globular fold, consisting of a hydrophobic core formed by five β-strands flanked by six α-helices that are connected through flexible loops (Figure 1.10A) [22, 53, 120]. The majority of differences between the A3 domain structures are predominantly found in loops 1, 3, and 7, which are near the active-site, as displayed in Figure 1.10B. This may reflect the observed differences in the intrinsic sequence preference (refer to 1.6.3.3 and 1.6.4), regulation of catalytic activity (refer to 1.5 and 1.6.4), physiological roles (refer to 1.4), and potential to oligomerise (refer to 1.5) [41, 122-123, 273].

The active-site is highly conserved between the A3s and related zinc-dependent deaminases, implying that the mechanism of cytosine deamination is conserved. The core of the active-site comprises a single zinc (Zn^{2+}) ion and a hallmark ZDD-motif: His-x-Glu-$x_{23–28}$-Pro-Cys-x_2-Cys (where x is any amino acid residue) localised at the N-terminal ends of the α2 and α3 helices. Important residues within the ZDD-motif, histidine (His) and two cysteines (Cys), coordinate Zn^{2+} ion into the active-site and the catalytic glutamic acid (Glu) is essential to mediate the reaction [63, 113, 142, 239] (discussed further in section 1.6.3.1).

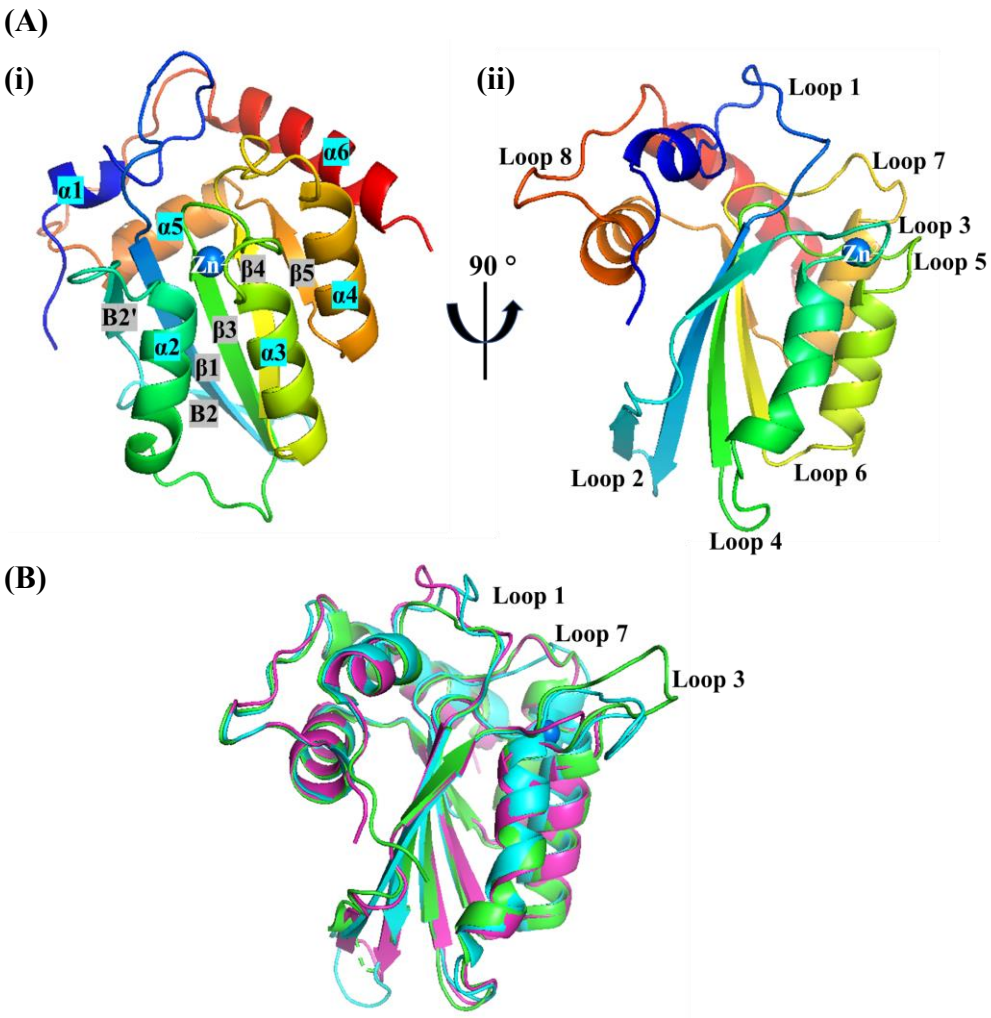


Figure 1.10 Crystal structures of the catalytic C-terminal domain of A3B.

(A) Crystal structure of $A3B_{CTD}$ PDB 5CQI ($A3B_{CTD}$-QM-ΔL3), (B) Comparisons between the $A3B_{CTD}$ (magenta, PDB; 5CQI), A3A (green, PDB; 4XXO), $A3G_{CTD}$ (cyan, PDB; 3IQS) structures. Cited from [240].

1.6.3 Biochemical insights of the A3 family

1.6.3.1 Cytosine deaminase mechanism

The cytosine deamination mechanism is based on early structural studies of bacterial and yeast CDAs [14, 120]. Initially, an active-site water molecule is deprotonated by Glu, resulting in a stabilised Zn^{2+}-hydroxide ion (OH^-). Deamination of the cytosine pyrimidine ring begins through the protonation of nitrogen 3 (N3) by Glu (coordinated by Zn^{2+}-OH^-), which permits the nucleophilic attack of carbon 4 (C4) of the cytosine by the hydroxide ion. This yields an unstable tetrahedral intermediate, which is followed by the successive protonation of the amino group (NH_2) and the release of ammonia (NH_3) forming uracil as the product [100, 188] (Figure 1.11).

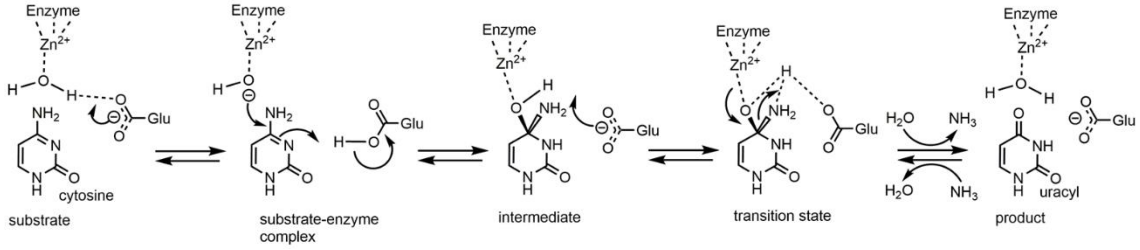


Figure 1.11 Model of the cytosine deaminase mechanism.

Courtesy of Assoc. Prof. Vyacheslav Filichev for producing this figure, cited from [14, 120].

As the ZDD-motif is conserved across the zinc-dependent cytosine deaminases, it is presumed that the mechanism described in Figure 1.11 is preserved for the A3 enzymes. This is consistent with this notion, upon mutagenesis of the residues (Glu, His or Cys) directly involved in the hydrolytic deamination mechanism the catalytic activity of the A3 enzymes is completely abolished [239]. In contrast to CDA, the A3 members are unable to catalyse free cytosine nucleosides or nucleotides, preferring cytosines within oligo- or polynucleotides (Figure 1.1) [15, 26, 48]. Early studies noted that the APOBEC family members (A4 excepted) contain two phenylalanine residues and an insert of four amino acids within the ZDD-motif (highlighted in grey in Figure 1.9) that are not present in the ancestral CDA. These residues were believed to permit the deamination of polynucleotide substrates and/or prevent deamination of single cytidines [2, 113, 184-185]. Furthermore, strands β3-5 in polynucleotide zinc-dependent deaminases are parallel to each other, unlike the antiparallel arrangement in bacterial and yeast CDAs. This arrangement probably leads to substrate specificity and may distinguish polynucleotide from non-polynucleotide deaminases [53, 60, 158, 205]. Finally, the A3 enzymes were found to require substrates with a minimum of three nucleotides [94, 180] to facilitate substrate binding and selectivity [239-240] (discussed further in section 1.6.4).

1.6.3.2 Nucleic acid preference of the A3 enzymes

Originally, it was assumed that the A3 family were RNA-deaminases like the A1 enzyme (refer to 1.3.1), based on similarities between the amino acid sequences [113]. However, the A3s were eventually found to be closer in sequence homology to AID, which is known to target DNA [95, 201]. Numerous experimental investigations found that the A3 family and AID strongly prefer the deamination of ssDNA over RNA and other modified oligonucleotides (such as methylated cytosines), but they are still capable of binding to either of the aforementioned substrates [40-41, 95, 98, 148, 155, 179-181, 220]. The main structural difference between DNA and RNA is the presence of a hydroxyl group on the C2'-

position of RNA, see Figure 1.12B. An earlier study reported that upon substitution of the target 2'-deoxycytidine with a 2'-ribose-nucleotide the catalytic activity of AID is severely compromised [180]. The slight change in the backbone sugar ring pucker, where DNA has a C2'-endo pucker (Figure 1.12A) and RNA has a C3'-endo pucker (Figure 1.12B), is thought to account for the observed selectivity. The conformation of RNA causes steric constraints in the enzyme's binding site, which cause alteration of the nucleotide's conformation. In contrast, the flexible rotational conformation found in 2'-deoxy-nucleotides better facilitates direct interaction with the enzyme promoting deamination of DNA over RNA (Figure 1.12A).

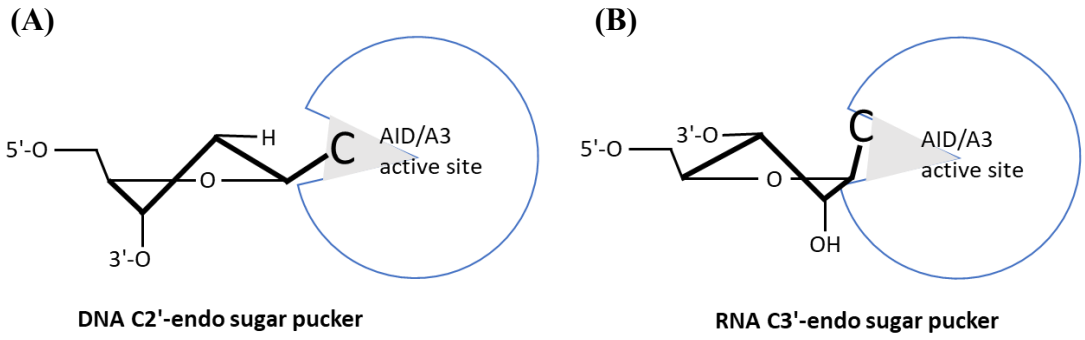


Figure 1.12 Major conformation of the 2'-sugar nucleotide in DNA and RNA.
Cited from [180].

This model was also supported by the recent molecular dynamics simulation of the A3B structure, suggesting that RNA substrates must undergo significant rearrangements with intermediate binding conformations to fit within the active-site [272]. Nonetheless, the precise mechanism of selectivity is not well understood. Furthermore, studies have shown that the A3 members and related family favour the deamination of 2'-deoxycytidines in ssDNA, with no evident activity on dsDNA [26, 51, 111, 252].

1.6.3.3 The A3 enzymes' nucleic acid deamination sequence

Studies have indicated that the A3s require ssDNA substrates with a minimum of three nucleotides, but to effectively bind and execute cytosine deamination five nucleotide bases are needed (note, length of oligonucleotide denoted by number followed by mer). As displayed in Figure 1.13, the target 2'-deoxycytidine (dC) is in position 0, whereas nucleotides 5'- of the target dC are given a negative number and nucleotides 3'- of dC are given a positive number. Three nucleotides must precede the target dC at the 5'-end (referred to as positions -1, -2 and -3), and a single nucleotide follows the target cytosine at the 3'-end (referred to as position +1) [94, 172, 180] (Figure 1.13).

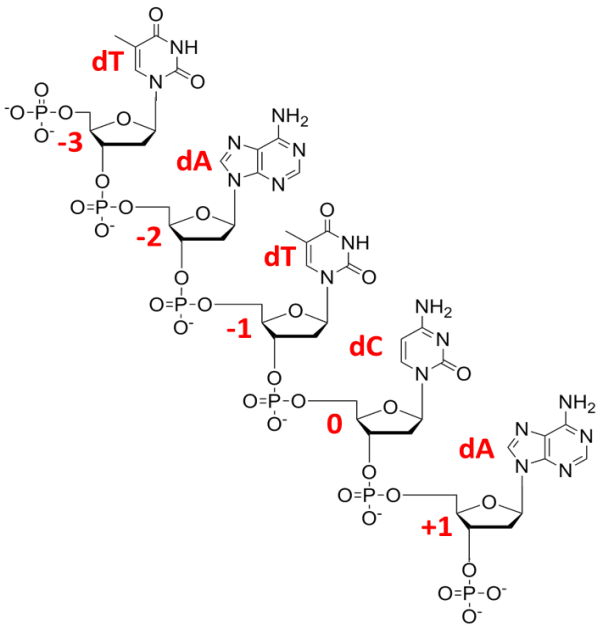


Figure 1.13 A3B$_{CTD}$ preferred ssDNA deamination sequence.
Sequence cited from [37]

The A3 enzymes' intrinsic deamination sequence not only dictates sequence recognition, but also the physiological role and signature markers of a variety of cancers (discussed in section 1.7.2). Based on mutation patterns and deep sequencing studies, the A3 members have been shown to deaminate the target dC based on the presence of a specific nucleotide base in the -1 position (refer to Figure 1.13) [11, 98, 123, 252, 255]. The majority of the A3 enzymes (A3A-F and A3H) prefer to deaminate cytosines preceded by a thymine (5'-T<u>C</u>, where the underlined C is the target), whereas A3G prefers to target cytosines preceded by another cytosine (5'-C<u>C</u>) as shown in Table 1.2. These enzymes have also been shown to be capable of targeting cytosines with other neighbouring nucleobases, but with lower efficiency [36-37, 239, 252].

Table 1.2 A3 family intrinsic dinucleotide deamination sequence preference (cited from [8])

Family member	Sequence specificity, 5'-3'
A3A	(T/C)<u>C</u>
A3B	(T/C)<u>C</u>
A3C	(T/C/G)<u>C</u>
A3DE	(A/T)<u>C</u>
A3F	(T/G)<u>C</u>
A3G	(C)<u>C</u>
A3H	(T/C)<u>C</u>

1.6.4 Structure of the inactive A3 domain in complex with ssDNA

Biochemical studies noted that ssDNA binds to the A3 enzymes by forming interactions with residues near the active-site [41, 53, 122-123, 207, 240, 242, 252, 273]. In 2016, the DNA-bound catalytically inactive A3A and the inactive $A3B_{CTD}$-A3A loop 1 chimera (termed $A3B_{CTD}$*) crystal structures were reported [239]. The conserved active-site Glu was mutated to Ala to destroy enzymatic activity. These structures demonstrated that their overall binding and mechanism selective to cytosine were essentially the same, so it is presumed that other A3 enzymes bind DNA in a similar manner. The DNA forms a kink at the target dC and preceding nucleobase (thymidine in the -1 position), adopting a U-shaped conformation [93, 239] (Figure 1.14A), which resembles the substrate-binding mode of a related adenosine deaminase (TadA) [158]. The 5'-TC nucleobases are flipped-over into the active-site, with respect to the phosphodiester backbone, whereas the nucleotides flanking this dinucleotide sequence are positioned away from the active-site. This base-flipping mechanism facilitates direct interactions with residues both surrounding and within the active-site, which together help anchor the target dC and access the catalytic glutamic acid in the active-site core [93, 239], as shown in Figure 1.14A and B.

Looking at the $A3B_{CTD}$*-ssDNA structure, the substrate binds in the active-site while interacting with loops 1, 3, and 7 (Figure 1.14B). Taking a closer look, the 5'-TC motif is initially stabilised through van der Waals interactions with W281 (loop 5). The -1 T fits into a hydrophobic pocket near the active-site formed between W281 (loop 5), Y313 and Y315 (loop 7) via van der Waals interactions. In addition, the -1 T is hydrogen bonded to D314 and Y315 (loop 7) within this pocket to further position it in place (see Figure 1.14C and D). The target dC is inserted deep into the active-site core, which contains the catalytic glutamic acid (E255 in helix α2) and the coordinated Zn^{2+} ion (zinc-coordinating residues, H253 (α2), C284 (loop 5), and C289 (loop 5)). The target dC undergoes extensive hydrogen-bond interactions with several residues (Y313 (loop 7), A254 (α2), S282 (loop 5), T214 (loop 1)) and is inserted between T214 and H253 to position it in the correct orientation for deamination (Figure 1.14C and D). These interactions are essentially the same, involving equivalent residues of the A3A (cyan, Figure 1.14B), as $A3B_{CTD}$ and A3A share 90 % sequence identity. However, the A3A-ssDNA structure additionally shows that His29 from loop 1 forms essential interactions with the nucleic acid substrate at multiple positions and stacks with Y132 to support the U-shaped

conformation of ssDNA [239]. This is believed to be similar in the wild-type $A3B_{CTD}$, where Arg211 seems to be equivalent to His29 of the A3A [239, 241, 272].

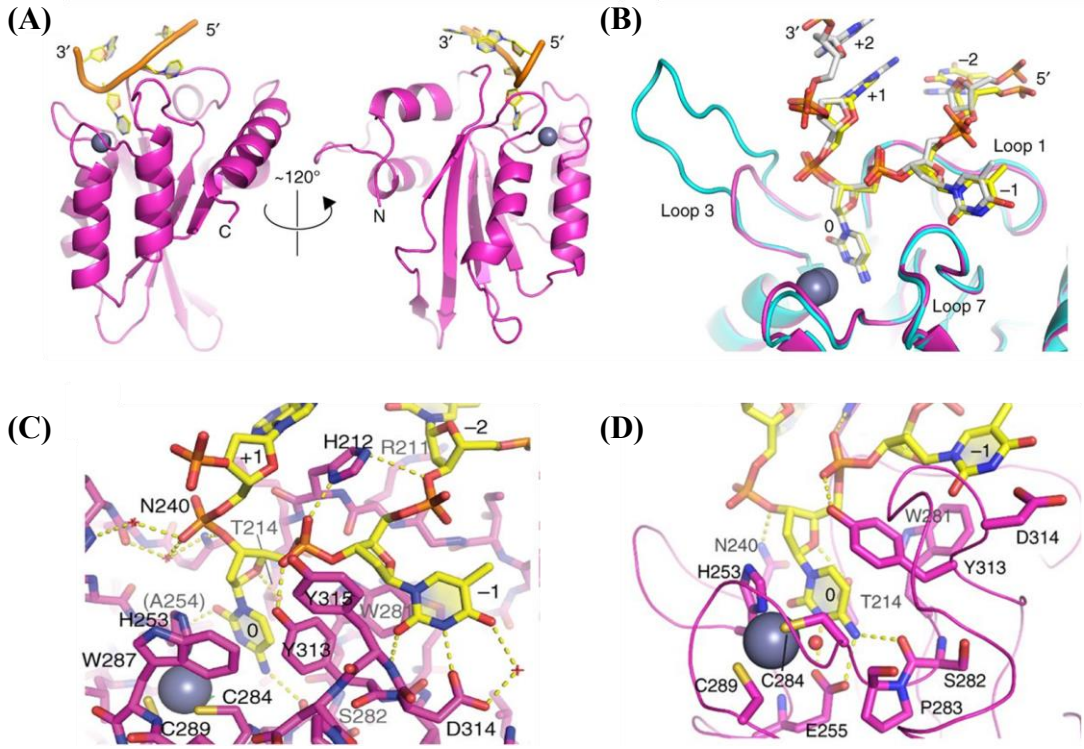


Figure 1.14 Crystal structure of inactive $A3B_{CTD}$* in complex with ssDNA.
Structure of the inactive $A3B_{CTD}$* ($A3B_{CTD}$-QM-ΔL3-A3ALoop1swap-E255A) in complex with a 5mer ssDNA sequence (5'-TTCAT) (PDB 5TD5 [239]). (A) $A3B_{CTD}$*-ssDNA structure showing ssDNA bound in a U-shaped conformation with the target dC and -1T flipped into the active-site. (B) Comparison of the active-site of the inactive $A3B_{CTD}$*-ssDNA (magenta) and inactive A3A-ssDNA (PDB, 5SWW [239]) shows comparable structures. (C) Interactions between the $A3B_{CTD}$* and the 5'-nucleotide of the oligonucleotide. (D) Interactions between the $A3B_{CTD}$* and the -1T and the target dC (position 0). Reprinted with permission from Springer Nature [239], copyright (2017).

Structure-based mutagenesis studies on A3A identified the specific residues required to dictate the A3 enzyme's dinucleotide sequence selectivity and mediate its catalytic activity [239]. As expected, mutating residues involved in cytosine deamination such as the catalytic glutamic acid and the zinc-coordinating histidine and two cysteines to alanine completely abolished the deamination of the target dC. A3A residues that interact with the target cytosine A71, S99, and Y130 (conserved in $A3B_{CTD}$ as A254, S282, and Y313, see Figure 1.14C and D) were also found to be essential in facilitating the deamination of the target dC. However, the other A3A cytosine contact with T31 (T214 for $A3B_{CTD}$, Figure 1.14C and D) maybe dispensable, as the mutation T31A does not compromise cytosine deamination. However, T31 is believed to be phosphorylated in order to regulate

the enzyme's catalytic activity [67, 243], consistent with the observed lack of activity upon the introduction of a negatively charged amino acid, T31D. On the other hand, mutating the A3A residues W98 (W281 in $A3B_{CTD}$), Y130 (Y313 in $A3B_{CTD}$), and Y132 (Y315 in $A3B_{CTD}$) (see Figure 1.14C and D) with alanine renders the enzyme catalytically inactive, but replacement with other aromatic amino acids restores function. Therefore, the aromatic nature of these residues is vital for the formation of the hydrophobic pocket to accommodate and stabilise the -1 nucleobase of the ssDNA, whereas interactions with residue D131 in loop 7 of the A3A (D314 in $A3B_{CTD}$, Figure 1.14C and D) influence the selectivity of the -1 nucleobase towards thymine [239]. This is consistent with studies where substituting loop 7 between the A3 enzymes causes a complete swap in the enzymes' dinucleotide deamination preference [41, 122-123, 166, 207, 239, 256, 273].

Although neighbouring nucleobases either side of the preferred dinucleotide local sequence (5'-TC) show weak enzyme preferences, as implied by deep deamination experiments, it is believed that they form surface interactions to sustain the U-shaped conformation [239-240]. In the substrate-bound A3A and $A3B_{CTD}$* structures, nucleobases in positions +1 and -2 form non-specific interactions with residues within loops 1 and 3 of the enzyme (Figure 1.14C), whereas nucleotides further away stack on top of one another through hydrophobic interactions (+2, +3 nucleobases stack on the +1 nucleobase, and -3 nucleobase stacks on the -2 nucleobase) to stabilise U-shaped conformation [239] (Figure 1.14A and B). This is consistent with the molecular dynamics simulations, where the flanking nucleotides at either end are not entirely fixed in place due to their limited interaction with the protein [93]. Together, the additive effect of these non-specific interactions aids with overall stability of the conformation of the substrate, which in turn supports binding of ssDNA to the A3 enzymes.

Comparison between the substrate bound and unbound A3A and $A3B_{CTD}$* structures, illustrates that substrate binding to the A3s' active-site requires rearrangements of loops surrounding the active site (loops 1, 3, and 7) (Figure 1.15). In particular, the unbound $A3B_{CTD}$ was shown to have a tightly closed active-site conformation stabilised largely by stacking interactions between loop 1 (Arg211) and loop 7 (Tyr315) [240] (Figure 1.15B i). To achieve a more open conformation, significant reorientation of these loops is believed to occur to facilitate binding of the ssDNA substrate, as displayed in Figure 1.15B ii [239]. It is thought that this closed conformation maybe a mechanism to regulate the catalytic activity of this enzyme in the nucleus to prevent off-target reactions, but the mechanism

behind this remains elusive due to the lack of information on wild-type full-length A3B structure [241, 272]. In comparison, the A3A structure in both the DNA-bound and -unbound forms has a more open conformation, as displayed in Figure 1.15A [239], consistent with having the highest catalytic activity and binding affinity towards DNA in comparison to the other A3 enzymes [36-37, 94, 135].

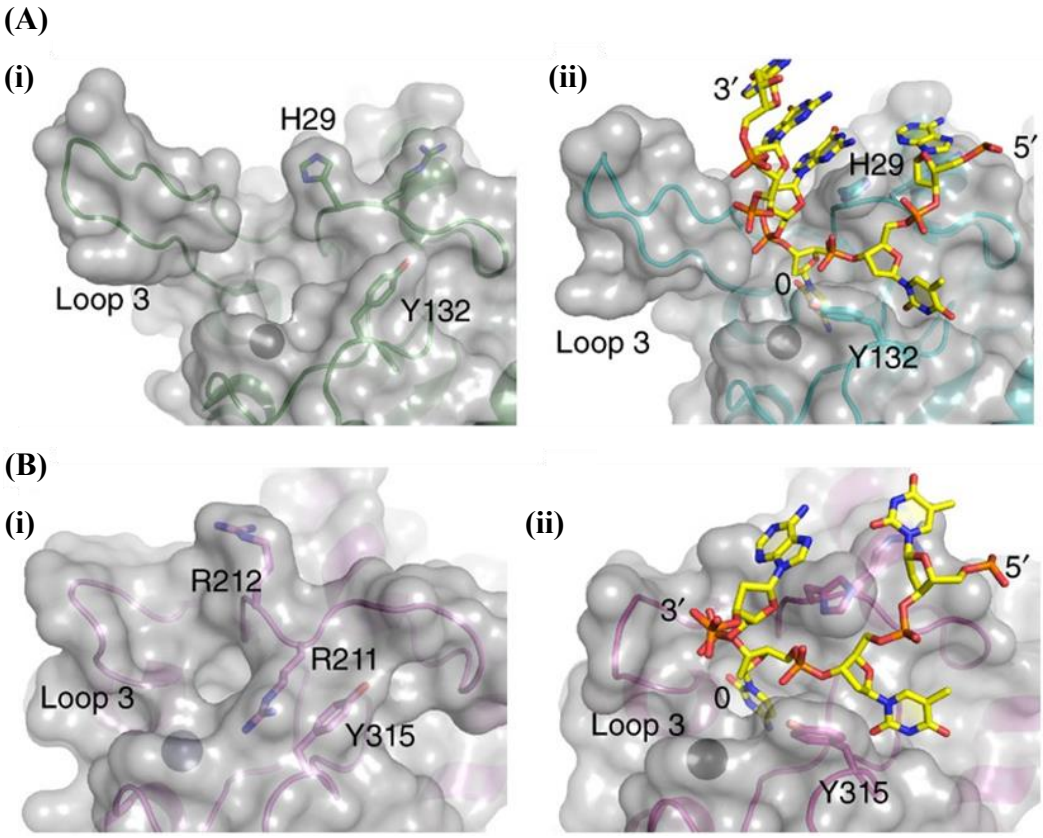


Figure 1.15 Comparisons between the ligand-free and ssDNA-bound active-sites of A3A and A3B*.

(A) Surface structure of the (i) ligand-free A3A structure (PDB, 4XXO [23]) and (ii) DNA-bound A3A structure (PDB, 5SWW [239]). (B) Surface structure of the (i) ligand-free $A3B_{CTD}$ structure (PDB, 5CQH) and the (ii) DNA-bound $A3B_{CTD}$* form (PDB, 5TD5 [239]). Reprinted with permission from Springer Nature [239], copyright (2017).

Switching loop 1 of $A3B_{CTD}$ to the corresponding loop 1 of A3A led to the enhancement of deaminase activity *in vitro* [37, 135, 240], consistent with a more open structural conformation [239] demonstrating the importance of loop 1 in binding ssDNA [294]. To accommodate binding of the nucleic acid substrate into the active site, interactions between residues in the loops surrounding the active site and the ssDNA cause a conformational change that further opens the active-site. This enables the dinucleotide sequence motif to establishment specific contacts in the active site, which then allows the

target dC to be positioned deep into the catalytic site where it can be deaminated by the conserved catalytic glutamic acid and the H_2O-coordinated Zn^{2+} ion [239].

1.7 A3s pathological consequences

Although the A3s functions in the innate immunity by protecting against retroviruses and pathogens, misregulation of their mutagenic action can have adverse pathological consequences.

1.7.1 Viral genome evolution

Despite the mechanistic action of Vif to inhibit the A3s' activity (described earlier in section 1.4.1), several studies found that G-to-A mutations of the HIV-1 genome persisted even in the presence of Vif [88-90, 174] (mechanism displayed in Figure 1.6). One study suggested that this may occur as A3F was more resistant to inhibition by Vif in comparison to A3G; however, this was only shown *in vitro* [154]. An alternate *in vivo* study noted that there was a high prevalence of defective Vif variants that led to incomplete or partially inhibited A3 enzymes, which would allow the A3 enzymes to restrict HIV-1 [246]. Apart from compromising the virus via lethal mutations, strong evidence has suggested that low levels of A3 enzymes that are not inhibited induce non-lethal mutations, which enables the virus to maintain infectivity and to contribute to its rapid evolution [96, 115-116, 219]. The HIV-1 retrovirus exploits the A3s' action by increasing mutation rates, which enable it to evade immunity clearance systems and develop antiretroviral drug resistance.

1.7.2 A3B and cancer evolution

In addition to targeting viral DNA and retrotransposons, several A3s have been implicated in mutations of genomic DNA, especially cancer cells [255]. Specific regions of the genome, mainly late replicating and heterochromatic regions, are more prone to mutations by A3 members as a result of ssDNA intermediates during stalled replication forks [34, 156]. DNA damage caused by cytosine deamination is usually restored in an error-free manner using the base excision repair enzyme, human UDG [251]. The uracil lesions obtained after misincorporation are excised using UDG resulting in abasic sites which can be cleaved by endonuclease to trigger repair mechanisms or the DNA is degraded [133, 252]. However, genomic uracil lesions and abasic site repair intermediates that fail to repair can become genotoxic and manifest as somatic mutations. These lesions and intermediates can promote a variety of mutations including C-to-T transition mutations, C-to-G transversion mutations, and other mutations following single-stranded DNA

breaks (SSB) or double-stranded DNA breaks (DSB) [46, 149, 261], displayed in Figure 1.16. The accumulation of these genomic DNA mutations, therefore, enables accelerated cancer evolution.

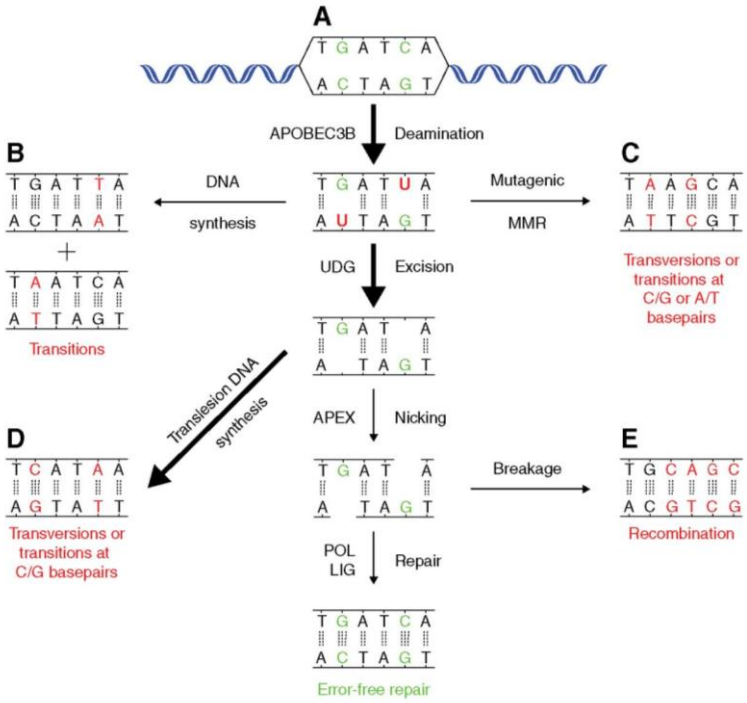


Figure 1.16 C-to-T mutations driven by A3B during repair of genomic uracils.
(A) Following deamination of genomic dC to dU by A3B, this event can be repaired by excising the uracils with UDG and triggering the base excision repair (BER) system (this includes APEX a BER protein, DNA polymerase termed Pol, and DNA ligase termed Lig). Misrepaired lesions or intermediates lead to a number of mutations. (B) Transition C-to-T mutations following DNA synthesis with a uracil in the template or an abasic site (DNA polymerases inserts dA opposite an abasic lesion). (C) Mismatch repair system (MMR) with a transition of transversion of C-to-T or A-to-T lesions. (D) Transversion C-to-G mutations thought to occur when DNA repair protein (REV1) inserts dC opposite the abasic site, then followed by DNA synthesis or repair of lesion. (E) Single-stranded DNA breaks (SSB) or double-stranded DNA break (DSB) result from phosphodiester backbone cleavage by APEX leading to recombination or further mutagenesis. Reprinted with permission from American Association for Cancer Research [149], copyright (2013).

Multiple studies noted that the overexpression and activity of A3A, A3B, A3D, and A3H were associated with cancer mutagenesis [1, 34, 95, 138, 140, 213, 256]. The C-to-T mutation signature in cancer was observed to occur within the hallmark 5'-TC(A/T) and 5'-TCG motifs, which are the preferred deamination motifs of both A3A and A3B. The given involvement of A3A and A3B in initiating cancer is yet not defined. However, A3B is thought to be the main contributor of genomic mutations as it is the only A3 member to

be constitutively localised in the nucleus of the cell [12, 34, 137-138, 147, 253]. A3A may facilitate genomic mutations and appear to be involved in some cancers [34-35, 190, 213], but, unlike A3B, is not found to be significantly overexpressed. On the other hand, A3B is found to be significantly overexpressed in a wide variety of cancers (including breast, bladder, cervix, lung, head, and neck) as a result of the high rate of mutations, suggesting that A3B is the predominant factor in cancer progression [1, 34-35, 39, 95, 101, 147, 149, 213].

The mechanisms by which the A3 enzymes are upregulated in cancer cells are not well understood, but it is thought that they may be triggered by DNA-based viruses. HBV has been reported to be associated with the carcinogenesis of hepatocellular carcinoma (HCC) as a result of the upregulation of A3B. The hepatitis viral protein (HBx) is encoded by the HBV genome; HBx stimulates HBV replication and hepatocyte survival. The A3 enzymes deaminate the HBV viral DNA generating a truncated form of HBx. This truncated HBx leads to neoplastic hepatocytes promoting HCC proliferation and survival [284], through yet unknown mechanisms. Moreover, A3B has been noted to be upregulated in early stages of cancer by another DNA-based virus, HPV. Infection by this virus drives A3B oncogenic off-targets [101, 191, 269, 275]. The elevated levels of A3B in tumours stimulate somatic mutations that promote cancer evolution, which in-turn causes cancer recurrence, metastasis, and resistance to anticancer therapies [146, 244].

1.8 Therapy by hypermutation or hypomutation

Viruses and cancer cells have developed methods to exploit the mutagenic action of A3s for their own gain, leading to severe consequences on human health [116, 259]. Therefore, modulating the A3s' action as therapeutic drugs may be used as a preventative measure against progression of viruses and cancers [195]. Two strategies have been proposed, therapy by hypermutation or therapy by hypomutation. In regard to HIV-1, therapy by hypermutation involves blocking Vif which essentially restores the action of the A3G/A3D/A3F/A3H, preventing HIV-1 replication. Several small-molecule inhibitors of Vif have been identified, but are in the initial stages of development and require structure-based information [7, 43, 159, 183, 200, 297]. On the other hand, therapy by hypomutation entails direct inhibition of A3G/A3D/A3F/A3H, thereby preventing the diversification of the virus and development of drug resistance [96]. Currently, low potency μM (3.9-21 μM) small-molecule inhibitors specific to A3G have been discovered [151, 194, 196]. However, small-molecule inhibitors of other A3 members, such as the mutagenic action of the A3B in cancer cells have not been reported. In regard to cancer, therapy by hypermutation

would entail enhancement of the A3Bs mutational activity to toxic levels by small molecules to destroy cancer cells, though this strategy may be too difficult. Therefore, therapy by hypomutation by directly inhibiting A3B would provide mechanism in preventing cancer evolution and development of drug resistance.

Approach and outline of this study

1.9 Approach of this study

The fast-paced evolution of cancer facilitated by the hyperactivity of A3B deamination that leads to tumour recurrence, drug resistance, and poorer survival rates among patients is a growing area of concern. Notably, it has been reported that an A3B deletion polymorphism was found in a substantial fraction of the human population [114]. Due to the non-essential nature of A3B indicated by these studies, its inhibition would unlikely cause a problem in normal cells. Therefore, the targeted inhibition of A3B, while retaining the activities of the other A3 enzymes involved in the innate immunity, may be used as a strategy to prevent cancer evolution and the development of drug resistances.

1.9.1 Rationale of substrate-like ssDNA inhibitors

The CDA replenishes pyrimidine nucleosides [296] but, several studies have noted that the CDA can also metabolise cytosine-based anticancer and antiviral drugs [57-58, 280]. This led to the development of several cytosine analogue nucleosides as potent inhibitors of CDA, which could be used as therapeutic conjuvants [6, 45, 277]. These cytosine analogues mimicked the transition state or intermediate structure of the substrate in an enzyme-catalysed deamination reaction (as displayed in cytosine deamination mechanism in Figure 1.11). The crystal structures showed that these inhibitors bound tightly into the active-site of CDA [14, 56, 281], consistent with the conversion of the activation energy required for catalysis into static binding energy [227] (as illustrated in Figure 1.17).

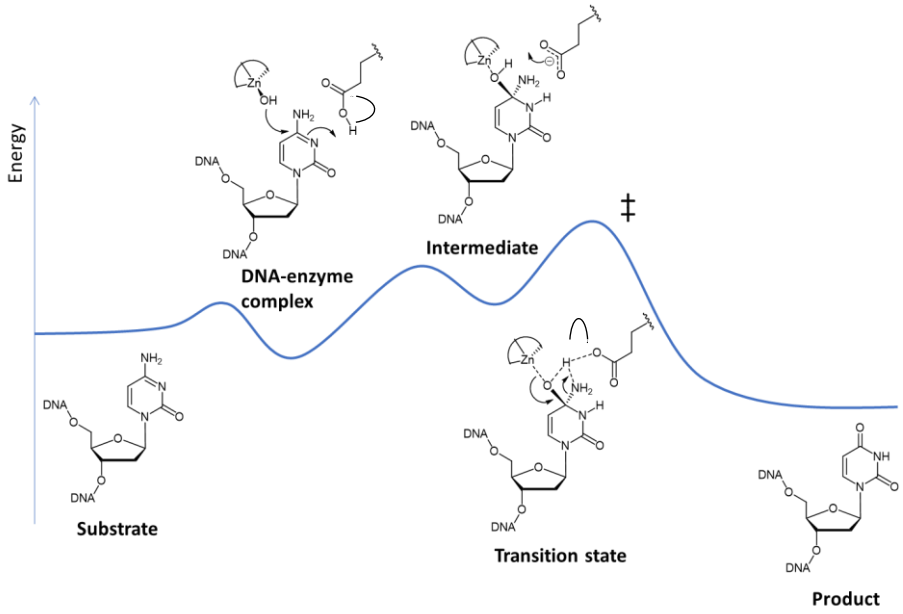


Figure 1.17 Scheme of cytosine deamination mechanism and associated conversion energy.
Courtesy of Dr. Maksim Kvach for producing this figure.

The A3 enzymes are believed to have a similar cytosine deamination mechanism as that of CDA [14, 120], as the catalytic domain between these enzymes is conserved (refer to 1.6.3.1). Therefore, the CDA cytosine analogue nucleoside inhibitors were presumed to inhibit the catalytic activity of the A3 enzymes in a similar manner. Initial assessment by our collaborators (Assoc. Prof. Daniel Harki's team, University of Minnesota) on the inhibition potential of some of these cytosine analogue nucleosides (tetrahydrouridine (THU), zebularine (rZ) Figure 1.18 [6, 45, 277], and a nucleotide 2'-deoxyzebularine (dZ)) on human A3A and human $A3B_{CTD}$ (expressed in human cells), indicated that these analogues did not inhibit the A3 enzymes [135]. This is coherent with A3 enzymes preferring to bind and deaminate cytosines only in ssDNA (nucleotides ≥ 5) [15, 26, 48], instead of free nucleosides as in CDA.

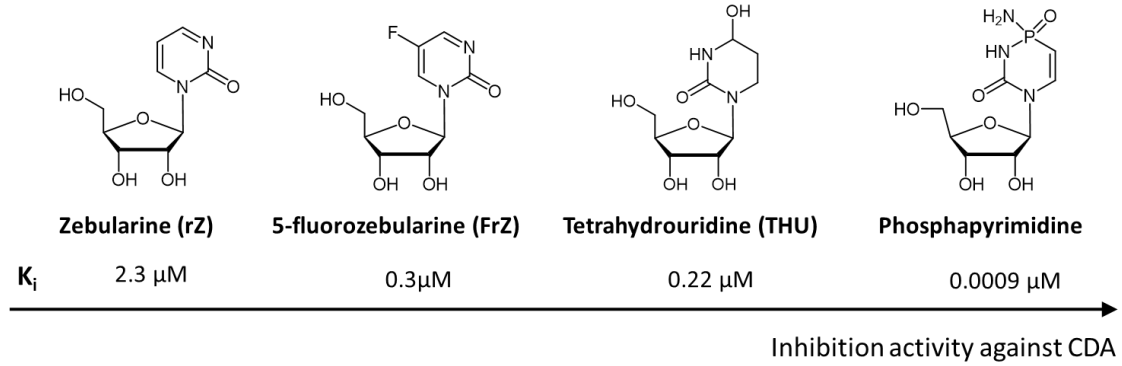
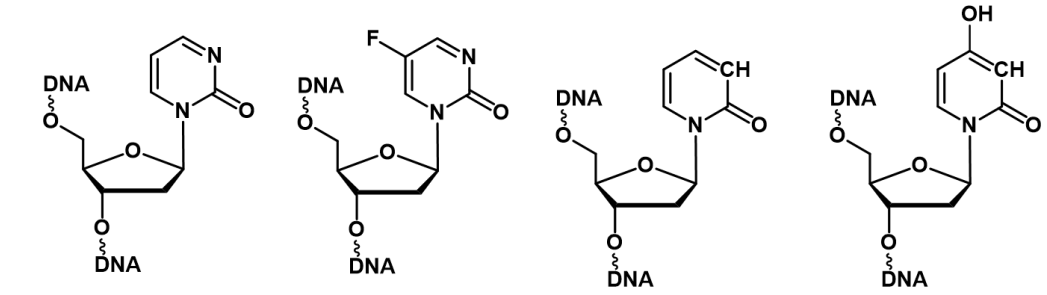


Figure 1.18 CDA cytosine analogue nucleoside inhibitors.

Courtesy of Dr. Maksim Kvach for producing this figure.

Therefore, our rationale to deliver these cytosine analogues into the A3s' active-site was to incorporate the deoxynucleoside form of some of these cytosine analogues into a short ssDNA sequence (ssDNA with 9-10 nucleotides chosen based on binding studies conducted by Dr. Stefan Harjes [135]). We embedded known CDA inhibitor cytosine analogue nucleosides, including 2'-deoxyzebularine (dZ), 2'-deoxy-5-fluorozebularine (5FdZ), 3-deaza-2'-deoxyzebularine (3dadZ), and 3-deaza-2'-deoxyuridine (3dadU), illustrated in Figure 1.19, in place of the target dC (position 0) within the A3B's preferred deamination motif (5'-TC, Table 1.2), to facilitate specific inhibition of A3B's catalytic activity. Modifying the chemical nature of the substrate ssDNA in this manner is believed to enhance its binding affinity towards the A3B by mimicking the transition state or intermediate of deamination (illustrated in Figure 1.17). The chemically modified ssDNA (termed modified-oligonucleotide) would therefore remain bound in the active site and thereby block the A3B's catalytic activity, thus, creating an A3B-specific ssDNA-based inhibitor.



2′-deoxyzebularine (dZ) | 2'-deoxy-5-fluorozebularine (5FdZ) | 3-deaza-2'-deoxyzebularine (3dadZ) | 3-deaza-2'-deoxyuridine (3dadU)

Figure 1.19 Design of chemically modified ssDNAs incorporated with cytosine analogues as potential inhibitors of A3B.

However, as A3B has an intrinsic deamination preference (5'-TC) similar to most A3s (except A3G) (see Table 1.2), these chemically modified inhibitors may not be selective towards A3B alone and inhibit other A3s. Apart from the A3's intrinsic deamination motif, several studies have demonstrated that the nucleotides flanking this motif also influence the enzyme's substrate recognition, and thereby deamination selectivity [37, 239-240]. Although the residues within the active site are highly conserved among A3 members, variations between the A3 members in loops 1, 3, and 7 that surround the active-site lead to differences in nucleotide contacts adjacent to the deamination motif [239-240]. This is consistent with the A3A and $A3B_{CTD}$ being 90 % identical in amino acid sequence, both enzymes prefer to deaminate cytosines in the same ssDNA sequences (5'-(T/A)TCA-motif), [239-240]. A study noted that these enzymes can differentially deaminate cytosines in other motifs with less efficiency, for example substrates with a CCC-motif [37]. A3A preferentially deaminates the dC at the 3'-end of the 5'-CC<u>C</u> motif as it prefers a purine in the +1 position, whereas A3B deaminates the dC at the 5'-end of the 5'-<u>C</u>CC motif as it prefers the presence of a thymine in the -1 position [8, 37, 239]. Therefore, our approach was to exploit this behaviour of A3A and A3B, by placing our tested cytosine analogue inhibitor (dZ) in the A3B's preferentially deaminated cytosine within the <u>C</u>CC-motif to determine if selective A3B inhibition is feasible (Table 5.1).

Our chemically modified-oligonucleotides were synthesised in-house by Dr. Maksim Kvach and Mr. Harikrishnan Mohana Kurup. First the cytosine analogue nucleosides and their phosphoramidites were synthesised, then they were combined into a short ssDNA sequence (experimental details and oligos are described in section 2.3.2).

1.9.2 Aim of this study

So far, no selective small-molecule inhibitors of A3B have been reported. Therefore, this research aimed at designing (see section 1.9.1) and evaluating substrate-like ssDNAs for their inhibitory potential on the A3B's catalytic activity. We also examined how the active A3 complexes interact with the chemically modified ssDNA using binding and structural studies. In understanding the molecular basis of interactions, the longer-term aim of this project is to provide a tool that inhibits the cancer-progressing actions of A3B and prevents the development of drug resistance against existing anticancer therapies.

(i) Express and purify $A3B_{CTD}$ for use in biophysical experiments.

Based on reported problems with expression and purification of the two-domain A3B, we opted to use the A3B catalytic C-terminal domain (termed $A3B_{CTD}$), as it is associated with the majority of the deaminase activity [78, 283]. For our experiments, we required the production of milligrams of soluble $A3B_{CTD}$ (described in section 2.2). Catalytically inactive $A3B_{CTD}$ variants were produced for binding studies to prevent deamination of dC in a substrate ssDNA, whereas catalytically active $A3B_{CTD}$ constructs were purified to determine the deamination of dC-containing ssDNA and evaluate the inhibitory potential of our modified ssDNAs (described above 1.9.1).

(ii) Evaluate the binding affinity of substrate ssDNAs and chemically modified ssDNAs on inactive $A3B_{CTD}$.

Using catalytically inactive $A3B_{CTD}$ constructs two methods, fluorescence-based thermal shift assay and isothermal titration calorimetry (ITC) (described in sections 2.4.1 and 2.4.2), were selected to understand the basis of interactions between A3 and ssDNA. These methods were used to screen the affinity of substrate ssDNAs for the enzyme and to indicate the inhibitory potential of the chemically modified ssDNAs to A3B.

(iii) Characterise the deamination potential of the active $A3B_{CTD}$ on substrate ssDNAs.

The established 1D ^{1}H-NMR-based activity assay [79, 94] (described in section 2.5.2) was selected to quantitatively characterise the deamination rate of catalytically active $A3B_{CTD}$ on substrate ssDNAs.

(iv) Evaluate the inhibitory potential of chemically modified ssDNAs against $A3B_{CTD}$.

The chemically modified ssDNAs were evaluated for their inhibitory potential on our active $A3B_{CTD}$ constructs by using a 1D ^{1}H-NMR-based inhibitor assay (described in section 2.5.3).

(v) Evaluate the ligand-free A3 and A3 in complex with ssDNA structures.
Two approaches were taken to elucidate the structure and structural features of the single domain A3B, as ligand-free $A3B_{CTD}$ and $A3B_{CTD}$ in complex with ssDNA. Solution-state NMR and chemical shift perturbations (described in sections 2.6.1 and 2.4.3) used the catalytically inactive $A3B_{CTD}$ with and without a substrate ssDNA. Small angle X-ray scattering (SAXS) used catalytically active $A3B_{CTD}$ with or without ssDNA-based inhibitors to enhance the binding affinity of DNA to the enzyme (described in sections 2.6.2). Furthermore, wild-type full-length A3G was used in SAXS studies to model the organisation of two-domain A3 with and without ssDNA-based inhibitors.

1.9.3 Thesis outline

Following the literature review and the approach of the study outlined in chapter 1, chapter 2 describes the methods used in this investigation. In chapters 3-6, I discuss the outcomes of this study. In chapter 3, I focus on understanding the molecular interactions between the purified inactive $A3B_{CTD}$ and ssDNAs through binding studies (i) and (ii), and also aimed at elucidating the solution-state NMR structures of the inactive $A3B_{CTD}$ and the $A3B_{CTD}$-ssDNA complex (i)(v). Following this chapter, the objective in chapter 4 centred on the assessment of the kinetic deamination of the target dC in the preferred ssDNA sequence (5'-TC, see section 1.9.1) by the catalytically active $A3B_{CTD}$ (iii). In addition, the inhibition potential of chemically modified ssDNAs containing cytosine analogue nucleosides was evaluated with the intention to obtain ssDNA-based inhibitors (iv). To further this study, in chapter 5, I explore whether selective inhibition of $A3B_{CTD}$ was feasible. Using the A3B discriminatory deamination of a suboptimal substrate (CCC-oligo), a cytosine analogue nucleoside replaced the preferred dC (dZCC-oligo as described in section 1.9.1) and oligo was evaluated for selective inhibition of $A3B_{CTD}$ (aims iii and (iv). In chapter 6, I present the SAXS models of $A3B_{CTD}$ and two-domain A3 in complex with our chemically modified ssDNA (i)(v). Then the overall conclusions and future directions are elaborated in chapter 7. Finally, all supporting figures, tables, and equations are displayed in the appendix section in chapter 8.

1.9.4 Hypothesis

Incorporating cytosine-analogue nucleosides in place of the substrate 2'-deoxycytidine in short ssDNA sequences will mimic the transition state of deamination and would inhibit the mutagenic action of A3B, by remaining bound in the active site.

2. Materials and Methods

2.1 General

2.1.1 Water, chemicals, and oligonucleotides

Unless specified otherwise, all water used for this work was milliQ, which was purified over two ion-exchange filters and two organic filters in a Barnstead NANOpure II system (Thermo scientific). MilliQ will be referred to as either water or H_2O throughout this thesis. Chemicals used were of the highest grade available. Unmodified oligonucleotides were commercially synthesised by IDT (Singapore).

2.1.2 Sterilisation

Media and buffers were sterilised by autoclaving at 121℃/15 p.s.i. for 20 minutes.

2.1.3 Antibiotics

Stock solutions of antibiotics were prepared as outlined in Table 2.1, filter sterilised through 0.22 μm Minisart filters (Sartorius Stedim), then stored as 1 mL aliquots at -20℃ for later use. Antibiotics were added to growth media or agar plates after autoclaving at working concentrations from stock to select for recombinant plasmid in *E. coli*.

Table 2.1 Preparation of the stock solutions of antibiotics

Antibiotic	Solvent	Stock solution (mg/mL)	Working concentration (μg/mL)
Kanamycin (Kan)	H_2O	100	100
Ampicillin (Amp)	H_2O	100	50

2.1.4 Luria-Broth (LB) medium and agar plates

Luria broth (LB) medium (Invitrogen) was made up at a concentration of 25 g/L with H_2O then sterilised by autoclaving. Once cooled to 50 °C, antibiotics were added to the medium at the working concentration described in Table 2.1. LB agar plates were made up by dissolving 25 g/L of LB and 15 g/L of agar bacteriological (Oxoid) in H_2O then autoclaved. As described antibiotics are added to the medium once cooled, and the medium is poured into plates to set.

2.2 Protein Expression and Purification

2.2.1 Plasmids

Expression of the full-length A3B in *E. coli* was reported to be unsuccessful [240]. As the catalytic C-terminal A3B domain ($A3B_{CTD}$) constitutes the majority of activity, $A3B_{CTD}$ was studied to understand the catalytic and structural entities of A3B. Recombinant expression plasmids that improve protein expression and purification encoding $A3B_{CTD}$ derivative constructs were gifted by Assoc. Prof. Hideki Aihara and Prof. Reuben Harris, University of Minnesota [239-240]. Plasmids utilised in this study are listed in Table 2.2, with modifications displayed in Figure 2.1.

Table 2.2 Recombinant plasmids (gifted by Assoc. Prof. Hideki Aihara and Prof. Reuben Harris, University of Minnesota [240])

Name	Description	Characteristics
$A3B_{CTD}$-QM-ΔL3	$A3B_{CTD}$ residues 187 – 378 with quadruple mutations F200S/W228S/L230K/F308K, loop 3 deleted (Ala242-Tyr250) and replaced with a serine residue; inserted into pET24a plasmid.	C-terminally His_6-tagged, Kanamycin resistant
$A3B_{CTD}$-QM-ΔL3-E255A	$A3B_{CTD}$ residues 187 – 378 with quadruple mutations F200S/W228S/L230K/F308K, loop 3 deleted (Ala242-Tyr250) and replaced with a serine residue, and amino acid substitution E255A; inserted into pET24a plasmid.	C-terminally His_6-tagged, Kanamycin resistant
$A3B_{CTD}$-QM-ΔL3-AL1swap - referred to as $A3B_{CTD}$-AL1 or A3A-mimic	$A3B_{CTD}$ residues 187 – 378 with quadruple mutations F200S/W228S/L230K/F308K, loop 3 deleted (Ala242-Tyr250) and replaced with a serine residue, and substitution of A3B loop 1 (DPLVLRRRQ) with A3A loop 1 (GIGRHK); inserted into pET24a plasmid.	C-terminally His_6-tagged, Kanamycin resistant
SUMO-$A3B_{CTD}$-DM	$A3B_{CTD}$ residues 187 – 378 with double mutations L230K/F308K; inserted into pE-SUMO plasmid	N-terminally His_6-tagged SUMO-fusion proteins, Ampicillin resistant
ULP1	SUMO protease encoded by ULP1 residues 1 – 621 inserted into pET19 plasmid	His_{10}-tagged SUMO protease, Ampicillin resistant

```
A3B_CTD                --------------PDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTWVLMDQHMGFLCN
A3B_CTD-DM             ------EILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTWVKMDQHMGFLCN
A3B_CTD-QM-ΔL3         ------EILRYLMDPDTFTSNFNNDPLVLRRRQTYLCYEVERLDNGTSVKMDQHMGFLCN
A3B_CTD-QM-ΔL3-E255A   ------EILRYLMDPDTFTSNFNNDPLVLRRRQTYLCYEVERLDNGTSVKMDQHMGFLCN
A3B_CTD-QM-ΔL3-AL1     ------EILRYLMDPDTFTSNFNNG---IGRHKTYLCYEVERLDNGTSVKMDQHMGFLCN
                             *  ** ****         *********  * * *   *  *** *

A3B_CTD                EAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQEN
A3B_CTD-DM             EAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQEN
A3B_CTD-QM-ΔL3         E--------SGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQEN
A3B_CTD-QM-ΔL3-E255A   E--------SGRHAALRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQEN
A3B_CTD-QM-ΔL3-AL1     E--------SGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQEN
                         **** * ***..*   **  * **** * *******  ** *   *.  *

A3B_CTD                THVRLRIFAARIYDY-DPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPW
A3B_CTD-DM             THVRLRIKAARIYDY-DPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPW
A3B_CTD-QM-ΔL3         THVRLRIKAARIYDY-DPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPW
A3B_CTD-QM-ΔL3-E255A   THVRLRIKAARIYDY-DPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPW
A3B_CTD-QM-ΔL3-AL1     THVRLRIKAARIYDY-DPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPW
                       ** * *  *****       *.*  *  *** .***** ** .******  ********

A3B_CTD                DGLEEHSQALSGRLRAILQNQGN
A3B_CTDDM              DGLEEHSQALSGRLRAILQ----
A3B_CTD-QM-ΔL3         DGLEEHSQALSGRLRAILQ----
A3B_CTD-QM-ΔL3-E255A   DGLEEHSQALSGRLRAILQ----
A3B_CTD-QM-ΔL3-AL1     DGLEEHSQALSGRLRAILQ----
                       ******** **********
```

Figure 2.1 A3B$_{CTD}$ recombinant variants utilised in this study.

* represents the same amino acid

2.2.2 Preparation of chemically competent bacterial cells

Bacterial strains listed in Table 2.3 used for transformations were first purchased from Lucigen. However, for cost and time efficiency these bacterial cells were made chemically competent using an in-lab established protocol. First, the desired *E. coli* strain was grown onto an LB plate at 37℃ overnight. Several colonies from the LB plates were inoculated into 250 mL SOB medium (0.5% Yeast Extract, 2% Tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$) and grown overnight at 19℃ to an OD of approximately 0.5-0.7 before being centrifuged at 2,500 × g for 10 minutes at 4℃. The supernatant was discarded, while the cell pellet was resuspended in 80 mL ice-cold transformation buffer (TB) solution (10 mM PIPES, 15 mM $CaCl_2$, 250 mM KCl, dissolved and adjusted to pH 6.7, 55 mM $MnCl_2$) and incubated on ice for 10 minutes. The resuspended cells were centrifuged and the supernatant discarded as described previously, after which the pellet was resuspended in 20 mL of ice-cold TB solution with 1.4 mL DMSO and stored in 100 µL aliquots at -80℃. Recombinant plasmids were first transformed into HI-control™ 10G (Lucigen) *E. coli* cells to propagate plasmids (see section 2.2.3). Following plasmid amplification, the plasmids were transformed into a bacterial protein expression system, BL21(DE3) (Lucigen) or C41(DE3)pLysS (Lucigen) (Table 2.3) (see section 2.2.4).

Table 2.3 *E. coli* strains

***E. coli* cell strain**	**Genotype**	**Source**	**Use in study**
HI-Control™ 10G	*mcrA* Δ(*mrr-hsdRMS-mcrBC*) *endA1 recA1* Φ80d*lacZ*ΔM15 Δ*lacX74 araD139* Δ(*ara,leu*)7697 *galU galK rpsL* (Str^R) *nupG* λ^- *tonA* Mini-F $lacI^{q1}$ ($Gent^R$)	Lucigen	Plasmid amplification
HI-Control™ BL21(DE3)	F –*ompT hsdSB (rB- mB-) gal dcm* (DE3)	Lucigen	Protein expression
OverExpress C41(DE3) pLysS	F –*ompT hsdSB (rB- mB-) gal dcm* (DE3) pLysS (CmR)	Lucigen	Expression toxic proteins

2.2.3 Plasmid amplification

The recombinant plasmid constructs (Table 2.2) were transformed into chemically competent *E. coli* cells to amplify and create plasmid stocks using an in-lab established protocol. First, 1 µL of plasmid was mixed into 50 µL ice-thawed chemically competent HI-control™ 10G (Lucigen) *E. coli* cells (Table 2.3), then incubated on ice for 30 minutes. The cells were placed in a TS1 Thermoshaker heating block (Biometra) at 42℃ for 30 seconds and then back on ice for 2 minutes. To allow the cells to recover from heat shock, 50 µL of LB media was added to the mix then incubated for 45 minutes at 37℃ with shaking at 225 rpm. The cell suspension was plated on LB agar plates (containing appropriate antibiotic) and incubated at 37℃ overnight for identification of successful transformants. For plasmid purification, a single colony from the overnight plates was then inoculated into 5 mL of LB-antibiotic medium and grown in a shaker overnight at 37℃. The culture was harvested by centrifugation and the plasmid was purified using a High Pure Plasmid Isolation Kit (Roche) according to the manufacturer's instructions. Following isolation the plasmid was verified (see 2.2.3.1 and 2.2.3.2) then stored at -20℃ for later use.

2.2.3.1 Agarose gel electrophoresis

Agarose gel electrophoresis was used to verify the size of the extracted plasmids. This method separates DNA fragments of varying sizes ranging from 100 bp to 25 kb. Agarose gels were prepared by dissolving the appropriate mass of agarose (AppliChem) into TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8.6) based on weight over volume (w/v) percentage solution, then heating the mixture until the agarose was

dissolved. The solution was cooled before pouring into a gel casting apparatus (Owl™ EasyCast™ Mini Gel Electrophoresis System, Thermo Scientific) and allowed to set at room temperature. DNA samples were mixed with a 6X loading dye (30% glycerol, 0.25% bromophenol blue) at a 1:1 (volume over volume, v/v) ratio before being loaded onto the gel and run at 100 V until the dye front migrated to the bottom of the gel. The agarose gels were then stained in ethidium bromide solution (0.5 µg/mL in H_2O) for 30 minutes before being transferred to H_2O for 15 minutes to remove residual stain. Then these gels were visualized using a BioRad GelDoc XR gel documentation system with Image Lab™ software.

2.2.3.2 Sequencing

The isolated plasmids were sequenced to verify that the plasmid contained the correct $A3B_{CTD}$ insert. Sequencing reactions were setup with 200 ng/µL plasmid along with 4 pmol of T7 forward primer (5' TAA TAC GAC TCA CTA TAG GG 3') in a total volume of 20 µL. Sequencing was performed by the Massey Genome Service (Massey University, Palmerston North), using a capillary ABI3730 Genetic Analyzer (Applied Biosystems Inc.). Nucleotide sequences were translated into amino acid sequences using ExPASy translate tool [83], then the amino acid sequences were aligned using protein BLAST® blastp suite [224].

2.2.4 Protein expression

Protein expression and purification of the $A3B_{CTD}$ constructs was mainly carried out as described previously [240]. The purified recombinant plasmids were transformed into chemically competent *E. coli* protein expressing strains (Table 2.3) by the heat shock method described earlier in section 2.2.3. In particular, recombinant plasmids, bearing genes for $A3B_{CTD}$-QM-ΔL3, $A3B_{CTD}$-QM-ΔL3-E255A, and the SUMO protease ULP1, were transformed into BL21(DE3) *E. coli* strain (Lucigen), whereas the more catalytically active derivatives $A3B_{CTD}$-QM-ΔL3-AL1swap and SUMO-$A3B_{CTD}$-DM were transformed into C41(DE3) pLysS *E. coli* strain (Lucigen) to minimise background expression of toxic proteins.

A single colony from the overnight transformed cells was inoculated into 10 mL of LB-antibiotic medium and grown overnight at 37℃ as a starter culture. The overnight starter culture was inoculated into 2 L of LB-antibiotic medium and grown in a shaker at 37℃ to an optical density at 600 nm (OD_{600}) of 0.3 for catalytically active $A3B_{CTD}$ enzymes,

and 0.6 for less active enzymes (mid-log growth phase). Cell culture OD_{600} is measured on a Smart Spec™ Plus Spectrophotometer (BioRad) using sterile LB medium as a blank. Once the optical density (OD) was established, the culture was cooled then supplemented with 100 µM final concentration of zinc chloride, to maintain Zn^{2+} within the A3B protein. Protein expression was induced by the addition of IPTG at a final concentration of 0.5 mM followed by incubation at 18℃ overnight. The cell culture was harvested via centrifugation at 4,000 × g for 30 minutes at 4℃ deriving a cell pellet which was frozen at -20℃ for further processing (see section 2.2.5).

2.2.5 Protein purification

The cell pellet was thawed then resuspended in lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 2 mM β-mercaptoethanol (β-ME)) containing one Complete™ EDTA-free protease inhibitor tablet (Roche) and 0.5 mg/mL of lysozyme. The resuspended cells were then lysed by sonication at 30% amplitude for 5 minutes (10 short bursts of 30 sec) using a sonicator probe (Qsonica). The whole cell extract was pelleted at 14,000 × g for 20 minutes to separate any unlysed cells, cellular debris, and insoluble protein.

Purification of the histidine-tagged proteins was conducted using two consecutive methods, immobilised metal affinity chromatography (IMAC), followed by size exclusion chromatography-fast protein liquid chromatography (SEC-FPLC) to ensure the extraction of pure protein. In IMAC, the histidine-tagged protein binds to a metal ion (such as Co^{2+}, Cu^{2+}, Ni^{2+}, and Zn^{2+}) that is covalently immobilised to a stationary support matrix (NTA) in the column. Non-specifically bound proteins are washed away from the column, the His-tagged proteins are eluted off the column by displacing this interaction using imidazole. Then SEC-FPLC further separates the proteins based on their size. Proteins flow through pores in the column resin, where large proteins pass by pores and elute first from the column, while small proteins penetrate pores leading to a delay in retention time and eventually elute from the column.

The clarified soluble cell lysate was filtered with a 0.8 µM filter before being loaded onto a pre-equilibrated with lysis buffer Ni^{2+}-NTA affinity Econo- column (BioRad) and subjected to gravity flow. The column was systematically washed with 20 mL of lysis buffer containing low concentration of imidazole (0 mM - 100 mM imidazole) to remove unbound proteins. The immobilised His-tagged protein was then eluted from the column by washing the resin with lysis buffers containing imidazole concentrations ranging

between 200 mM - 500 mM (dependent on protein construct). Fractions collected were then analysed by tricine SDS-PAGE gel (refer to 2.2.7.2), and the eluate was subsequently dialysed in high salt pH 7.4 buffer (50 mM Tris-HCl pH 7.4, 0.5 M NaCl, 2 mM β-ME) over 48 hours at 4 °C in a 10 kDa MWCO dialysis bag to remove imidazole. The dialysed protein was concentrated down to 0.5 mL in a 10 kDa cut-off Vivaspin 20 concentrator (GE Healthcare) then loaded onto a Superdex 75 10/300 GL column (GE Healthcare) using ÄKTA protein purification system (GE Healthcare) monitored using UV280 nm. The protein of interest was eluted using SEC-FPLC in SEC buffer (50 mM Tris-HCl pH 7.4, 0.5 M NaCl, and 2 mM β-ME) at a flow rate of 0.5 mL/min. Fractions collected were then visualised by SDS-PAGE gels (refer to 2.2.7.2) and the fractions containing the purified protein of interest were concentrated (Vivaspin 20 kDa MWCO) and stored in aliquots at -80℃ for later work.

2.2.5.1 Cleavage of fusion protein tag

The SUMO-tagged fusion protein (SUMO-A3B$_{CTD}$-DM) and SUMO protease (ULP1) were purified as described in section 2.2.5. Following purification, the SUMO-tag of the purified SUMO-A3B$_{CTD}$-DM was cleaved off in a similar manner to a previously described study [152]. In brief, the purified SUMO-A3B$_{CTD}$-DM protein was incubated with the purified SUMO protease (ULP1) at a ratio of 1:5 (fusion protein to protease) overnight at 4℃. After cleavage of the fusion protein, the sample was passed through a Ni^{2+}-NTA affinity column, where the His-tagged SUMO and His-tagged SUMO protease bound to the column, and the A3B$_{CTD}$-DM eluted off the column. The eluate was concentrated (Vivaspin 20 kDa MWCO) and stored in aliquots at -80℃ for further work.

2.2.6 Expression and purification of uniform isotopically labelled proteins

Uniform isotopically labelled proteins provide visible atomic nuclei used in multidimensional NMR experiments of proteins (further described in section 2.6). Proteins naturally contain ^{1}H isotope at a high abundance, but other isotopes are artificially enriched by using minimal medium containing ^{13}C as the sole carbon source or ^{15}N as the sole nitrogen source; both isotopes have a nuclear spin of ½, allowing a nuclear magnetic moment. Isotopically labelled proteins were expressed in M9 minimal medium in a similar manner to previous studies [5, 53, 102, 164]. Components of M9 minimal medium were first prepared as separate sterile stock solutions, outlined in Table 2.4 to Table 2.6. M9 minimal medium was then prepared, using either ^{15}N isotope or both ^{15}N and ^{13}C isotopes, as outlined in Table 2.7, then filter sterilised and adjusted with NaOH

to pH of 7.5 to maintain buffering capacity [5]. Initially, a transformed starter culture (see to 2.2.4) grown to high densities in LB medium is pelleted then exchanged into 5 mL of isotopically labelled M9 minimal medium and incubated for an hour at 37℃ to adjust to medium. This culture is then inoculated into the larger 2L of labelled M9 medium (supplemented with either ^{15}N or both ^{15}N and ^{13}C) and agitated at 37℃ in a shaker until the OD_{600} is around 0.4. Once the optical density was established the culture was cooled and its pH was adjusted to pH 7.5 using 1 M NaOH to obtain higher protein yields [5]. Then $ZnCl_2$ and IPTG were added at a final concentration of 100 µM and 0.5 mM to the culture respectively, and incubated overnight at 18℃. The culture was harvested and the labelled-proteins were purified using Ni^{2+}-NTA affinity chromatography and SEC-FPLC as described earlier in section 2.2.5. The purified isotopically labelled proteins were then stored at -80℃ for future work.

Table 2.4: M9 salts (5X)

Component	**Amount (g/L)**
$Na_2HPO_4.6H_2O$	34
$KH_2PO_4.H_2O$	15
NaCl	2.5
	Made up with H_2O to 1 L

Table 2.5: Trace element solution (1000x)

Component	**per 100 mL**
$CaCl_2.2H_2O$	0.60 g
$FeSO_4.7H_2O$	0.60 g
$MnCl_2.4H_2O$	0.12 g
$CoCl_2.6H_2O$	0.08 g
$ZnSO_4.7H_2O$	0.07 g
$CuCl_2.2H_2O$	0.03 g
H_3BO_3	2 mg
$(NH_4)_6Mo_7O_{24}.4H_2O$	25 mg
Na_2EDTA	0.50 g
	Made up with H_2O to 100 mL

Table 2.6: Vitamin solution (5 mg/mL) (stored at 4°C)

Component	per 100 mL
$C_{12}H_{18}Cl_2N_4OS$ (Thiamine hydrochloride)	0.50 g
$C_6H_5NO_2$ (Nicotinic acid)	0.50 g

Table 2.7: Isotopically labelled M9 minimal medium composition (^{15}N,^{13}C-labelled).

Component	Amount
M9 salts (5X) (see Table 2.4)	400 mL
Vitamin solution (see Table 2.6)	12 mL
Kanamycin (100 mg/mL)	700 µL
Trace elements (1000x)	2 mL
1 M $MgSO_4$	6 mL
$(^{15}NH_4)_2SO_4$	2 g
^{13}C D-glucose *****	4 g
	Made up with H_2O to 2 L

* note for ^{15}N-labelled proteins standard D-glucose was used.

2.2.7 Protein quantification

2.2.7.1 Protein Concentration

Protein concentrations were determined using the Bradford protein assay [25]. Albumin from chicken egg white (Sigma-Aldrich) was used to prepare a standard curve. Briefly, in a plastic cuvette 20 µL of standard or protein sample was added to 1 mL of Bradford reagent (Table 2.8). Samples were incubated for 5 minutes then their absorbances were measured on a Smart Spec™ Plus Spectrophotometer (BioRad) at an absorbance of 595 nm with the appropriate buffer used to blank the instrument. Protein concentration was calculated based on the standard curve, in which absorbance was plotted against varying concentrations of the albumin protein. Protein concentrations were also measured using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific) at an absorbance of 280 nm (A_{280}) with appropriate buffers used as a blank. Protein concentrations were calculated using Beers Law with the appropriate extinction coefficient calculated by ExPASy-ProtParam [84].

Table 2.8: Preparation of Bradford reagent

Component	Amount
85 % Phosphoric acid	100 mL
Brilliant blue G-250	100 mg
	Made up with H_2O to 1 L

2.2.7.2 Tricine SDS-PAGE Electrophoresis

To visualise proteins and assess molecular mass between 1-100 kDa Tricine SDS-PAGE gel electrophoresis was utilised as described in a previous study [225]. Tricine-SDS PAGE gels were utilised to resolve small proteins less than 30 kDa. Initially, buffers were prepared as outlined in Table 2.9. Gels were prepared using a Mini-Protean II Electrophoresis Cell System (Bio-Rad) at a thickness of 0.75 mm by first preparing the separating gel as outlined in Table 2.10, and casting in a gel cassette. Once the separating gel polymerised the stacking gel (also outlined in Table 2.10) was overlaid on top of it. Samples were mixed with the 2X SDS-treatment buffer at a 1:1 ratio (

Table 2.11), then were boiled for 5 minutes in a waterbath before being loaded onto the gel. Electrophoresis was conducted with 1X cathode buffer in the upper tank and 1X anode buffer in the lower tank (Table 2.9) and run at 100 V until the dye front migrated to the bottom of the gel. The Tricine SDS-PAGE gels were then stained with Coomassie blue stain (0.25% Coomassie Brilliant Blue R-250, 40% methanol, and 7% acetic acid) for 30 minutes (or overnight) with gentle agitation, followed by replacing the stain solution with destain solution (40% methanol, 7% acetic acid) and incubating overnight. The gel was placed into H_2O then imaged using BioRad GelDoc XR gel documentation system with Image Lab™ software.

Table 2.9 Preparation of Tricine SDS-PAGE buffers.

Component	Anode buffer (10X) pH 8.9	Cathode buffer (10X)	Gel buffer (3X) pH 8.45
Tris HCl	1 M	1 M	3 M
Tricine	-	1M	-
SDS	-	1 %	0.3%

Table 2.10 Preparation of Tricine SDS-PAGE gel.

Component	4% Stacking gel	16% Separating gel
Acrylamide (49.5%)/Bis-Acrylamide (3%)	0.5 mL	5 mL
Gel buffer (3X)	1.5 mL	5 mL
Glycerol	-	1.5 mL
H_2O	4 mL	3.45 mL
10% w/v ammonium persulfate	45 µL	50 µL
TEMED	4.5 µL	5 µL
Total	6 mL (1 mL used)	15 mL (3.5 mL used)

Table 2.11 Preparation of 2X SDS-treatment buffer

Component	Final concentration
1 M Tris-HCl, pH 6.8	100 mM
SDS	4 % (w/v)
Glycerol	10 % v/v
DTT	200 mM
Bromophenol Blue	60 µM

2.2.7.3 In-gel tryptic digest mass spectrometry

To confirm the identity of the eluted proteins in-gel tryptic digest mass spectrometry was employed using an in-lab established protocol. To obtain enhanced staining of the gel, colloidal Coomassie was utilised due to its high sensitivity to protein and reduced background staining in contrast to standard Coomassie blue stain (described 2.2.7.2). First, the SDS-PAGE gel was fixed in solution (40 % methanol, 10 % acetic acid), for an hour, then the gel was washed with H_2O, before being stained overnight in diluted colloidal Coomassie stain (working concentration made up of 4 parts stock to 1 part methanol) (Table 2.12). The gel was then washed in a 10 % acetic acid solution to remove any residual stain and readied for in-gel tryptic digest.

Table 2.12 Colloidal Coomassie stock solution.

Components	Concentration (w/v)
Coomassie Brilliant Blue G250	0.1%
Phosphoric acid	2%
Ammonium sulfate	10%
	Made up to 500 mL with H_2O (working solution 4:1, stock to methanol)

In-gel tryptic digest begins by excising the protein band of interest from the colloidal Coomassie-stained gel, then cutting it into smaller fragments and placing them in a 1.5 mL Eppendorf tubes. The gel fragments are first destained by incubation with 300 µL fresh 50 mM ammonium bicarbonate (ABC) solution at 45℃. Once the gel fragments are colourless the solution is decanted. Second, 300 µL of 80% aq. acetonitrile solution (ACN) is added then incubated for 1-2 minutes at room temperature, before discarding the solution. The gel fragments are dehydrated under a vacuum using a SpeedVac centrifugal evaporator (Thermo Scientific). This is followed by incubation with 50 µL of fresh reducing agent (10 mM DTT in 50 mM ABC) at 37℃ for 1 hour, then the reducing agent is removed. The gel fragments are washed with 50 mM ABC solution, then 80% aq. ACN solution, and dried using the SpeedVac as previously described. The gel fragments are incubated with 50 µL of fresh alkylation solution (20 mM iodoacetamide in 50 mM ABC) for 20 minutes in the dark at room temperature. After the alkylation solution is discarded, the gel pieces are again washed in 50 mM ABC, twice with 80% aq. ACN, and dried using a SpeedVac as described earlier.

Gel fragments are digested with TPCK-treated trypsin (Proteomics Grade T6567, Sigma) (20 ng/µL trypsin in 50 mM ABC solution) then incubated for 10 minutes on ice, before discarding the solution. This is followed by the addition of 30 µL of 50 mM ABC solution and incubation overnight at 37℃. After the incubation period, the samples are sonicated for 2 minutes and the supernatant is collected in a new Eppendorf tube. The gel fragments (in the original tube) can be further extracted; primarily 60 µL of extraction buffer 1 (5% formic acid in 80% aq. ACN) is added to the tube then sonicated for 2 minutes before pooling the supernatant with the previous collection; this is repeated with 60 µL of extraction buffer 2 (0.1% formic acid in 80% aq. ACN) and the supernatant is combined

with the previous collection. The combined supernatant was concentrated down to 20 µL then transferred to HPLC sample vials ready for mass spectrometry. Mass spectrometry and analysis were then performed by Mr. Trevor Loo, by processing samples on an Agilent 6520 Q-TOF mass spectrometer (Agilent Technologies) and utilising Agilent MassHunter Workstation Qualitative Analysis software version B.03.01 (Agilent Technologies, Santa Clara, CA, USA) and Mascot servers (version 2.4.1) housed in Walter and Eliza Hall Institute (Melbourne, Australia).

2.3 Single-stranded DNA substrates and potential inhibitors

2.3.1 ssDNA substrates

A3 enzymes deaminate cytosine residues in ssDNA with an intrinsic sequence preference. Several ssDNA sequences with 9-10 nucleotides were designed based on previously known sequences that interact with $A3B_{CTD}$ [36-37, 47, 78, 94, 157, 172]. The ssDNA oligonucleotides were commercially synthesised (IDT) containing 2'-deoxycytidine (dC) or 2'-deoxyuridine (dU) (substrate or product, respectively) (listed Table 2.13). These oligonucleotides were utilised in various experiments including thermal shift assay, isothermal titration calorimetry, and NMR titrations for relative binding capability to A3 variants (see section 2.4.12.4.2, and 2.4.3), and used in NMR-based activity assay (section 2.5.1).

Table 2.13 ssDNA oligonucleotides commercially synthesised (IDT) used in this study

Name	Sequence 5' → 3'
9merC-oligo (also termed TCA-oligo)	5'-ATTTCATTT
9merU-oligo	5'-ATTTUATTT
9ATC-oligo	5'-AAATCAAAA
9ATU-oligo	5'-AAATUAAAA
A9C-oligo	5'-AATTCAAAA
A9U-oligo	5'-AATTUAAAA
CCC-oligo	5'-ATTCCCAATT
CCU-oligo	5'-ATTCCUAATT
CUC-oligo	5'-ATTCUCAATT
UCC-oligo	5'-ATTUCCAATT
CUU-oligo	5'-ATTCUUAATT
UUU-oligo	5'-ATTUUUAATT
TC-3' 6-FAM-oligo	5'-ATTATTATTATTCAAATGGAT TTATTTATTTATTTATTTATTT-fluorescein

2.3.2 Chemically modified ssDNA

Towards the goal of blocking the progression of cancer evolution by A3B activities [34-35, 68, 146, 244], chemically modified ssDNAs were designed and then evaluated for their inhibitory potential against the catalytic activity of $A3B_{CTD}$. Known CDA inhibitor cytosine-analogue nucleosides, including 2'-deoxyzebularine (dZ), 2'-deoxy-5-fluorozebularine (5FdZ), 3-deaza-2'-deoxyzebularine (3dadZ), and 3-deaza-2'-deoxyuridine (3dadU), were incorporated in place of the target dC (position 0) in the A3B- and A3G-preferred ssDNA sequences (see Table 2.14). These analogues are believed to mimic the transition state of cytosine deamination, which increases the binding affinity owing to the conversion of the activation energy to static binding energy as no catalysis occurs [227]. The cytosine analogue nucleosides, their phosphoramidites, and ssDNA (oligos) containing these cytosine-analogue nucleosides were synthesised in-house by Dr. Maksim Kvach and Mr. Harikrishnan Mohana Kurup. Descriptions of the synthesis of these chemically modified oligos as well as data about composition and purity of modified oligos used in this work are available on-line in the supplementary information of our published articles (DOI:10.1021/acs.biochem.8b00858 [135], DOI: 10.1039/c9ob01781j [8], and DOI: https://doi.org/10.1002/cbic.201900505 [136]). These sequences were then tested for their inhibitory potential and binding affinity using isothermal titration calorimetry (section 2.4.2), NMR-based inhibition assay (section 2.5.3), and then examined using SAXS (section 2.6.2).

Table 2.14 ssDNA oligonucleotide synthesised in-house used in this study

Name	Sequence 5' → 3'	Modification
dZ-oligo	5'-ATTT**dZ**ATTT	dZ; 2'-deoxyzebularine
5FdZ-oligo	5'-ATTT**5FdZ**ATTT	5FdZ; 2'-deoxy-5-fluorozebularine
3dadZ-oligo	5'-ATTT**3dadZ**ATTT	3dadZ; 3-deaza-2'-deoxyzebularine
3dadU-oligo	5'-ATTT**3dadU**ATTT	3dadU; 3-deaza-2'-deoxyuridine
dZCC-oligo	5'-ATT**dZ**CCAATT	dZ
CCdZ-oligo	5'-ATTCC**dZ**AATT	dZ
$CCdZT_{30}$-oligo (termed CCdZ-40mer)	5'-ATTCC**dZ**AATTT_{30}	dZ, poly T tail at 3'-end

2.3.3 Nucleic acid concentration

Nucleic acid (plasmids, and ssDNA) concentrations were evaluated on a Nanodrop ND-1000 spectrophotometer (Thermo Scientific) at both 260 and 280 nm. Extinction coefficients (ε constant) were calculated using IDT OligoAnalyzer tool [197], then fed through ND-1000 software V3.1.0 (ThermoFisher) to calculate concentrations using the absorbance measured.

2.4 Binding studies

Binding interactions between the $A3B_{CTD}$ and ssDNA were studied using a combination of thermal shift assays and NMR titrations. Binding assays were typically conducted using a catalytically inactive $A3B_{CTD}$ variant to evade deamination of the cytosine in the sequence. Otherwise, binding of active $A3B_{CTD}$ variants to ssDNA inhibitors was assessed using isothermal titration calorimetry.

2.4.1 Fluorescence-based thermal shift assay

To investigate the binding of ssDNA to $A3B_{CTD}$, a fluorescence-based thermal shift assay was performed as described in a previous study [112]. This assay utilises a fluorescent dye to monitor the changes in the protein's thermal stability upon binding of a ligand. Initially, the purified $A3B_{CTD}$ protein was appropriately diluted in a 50 mM citrate-phosphate buffer (50 mM citrate-phosphate pH 5.5, 200 mM NaCl, 2 mM β-ME, pH 5.5, 200 μM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS)). In a white Low-Profile 96-Well plate (Roche) assay reactions were setup in a total volume of 25 μL containing 20 μM $A3B_{CTD}$ protein, 100 μM ssDNA oligonucleotide, mixed with a final concentration of 10X SYPRO® orange dye (BioRad). The plate was sealed with optical tape, shaken, and centrifuged before being subjected to temperature cycling. Thermal scanning (20℃ to 95℃ at 0.6℃/min) was performed using a real-time PCR setup on a LightCycler 480 instrument II (Roche) with fluorescence emission spectra recorded with combinations of excitation and emission filters (483–610 and 483–568 nm, respectively). Fluorescent intensity was normalised to percentage, then fit to curves to derive the melting temperature (T_m), which occurs at the midpoint of the unfolding transition. Binding of a ligand (ssDNA) to protein was evaluated by the change in melting temperature (ΔT_m), by comparing T_m in the presence of a ligand versus the absence of a ligand (buffer). Experimental replicates (n = 3-8) were performed, which were analysed using a Q-test for the identification and rejection of outliers based on a 95 % confidence interval.

2.4.2 Isothermal titration calorimetry (ITC)

ITC experiments were conducted to characterise the thermodynamic parameters of direct interactions between a protein and target ligand based on a previous study [76]. ITC relies on monitoring the changes in heat that is either released or absorbed during a binding interaction in solution. Firstly, the oligonucleotide and the $A3B_{CTD}$ protein were dialysed separately in a 50 mM citrate-phosphate pH 5.5 buffer (50 mM citrate-phosphate buffer, 200 mM NaCl, 2 mM β-ME, pH 5.5) for 2-3 days. ITC experiments were performed using a Micro-Cal ITC200 (Malvern Instruments) at 25℃, where the sample cell was loaded with 100 μM $A3B_{CTD}$ protein and the syringe loaded with 300 μM ssDNA. The protein was titrated with oligonucleotide over an 18 step injection of 2.0 μL each (plus an initial injection with reduced volume of 0.4 μL to prevent dilution of the DNA in the syringe due to the long wait before the start of the experiment). Analysis of ITC data was performed using Origin 7 ITC200 software where the raw data were fitted to a one-site binding model to derive parameters including enthalpy (ΔH), entropy (ΔS), equilibrium binding association constant (K_a), and stoichiometry (N). The equilibrium dissociation binding constant (K_d) can then be derived as $K_d=1/K_a$, from which the Gibbs free energy ($\Delta G^o = \Delta H^o - T\Delta S^o = -RT \ln (K_a/Q^o)$ can be calculated and combined with the measured ΔH^o to give ΔS^o, where T is temperature of 298 K).

2.4.3 NMR titration

NMR titration experiments were used to evaluate the binding site of ssDNA to the $A3B_{CTD}$ using a protocol previously described in several studies [92-94]. 2D-^{15}N HSQC NMR experiments (as detailed in 2.6.1 and Table 8.1) were utilised by examining the chemical shift changes of amino acids upon titration with ssDNA. Initially, ^{15}N-labelled $A3B_{CTD}$ (see 2.2.6) and the oligonucleotide were first dialysed separately in 50 mM citrate-phosphate pH 5.5 buffer (50 mM citrate-phosphate buffer, 200 mM NaCl, and 2 mM β-ME, pH 5.5). A ^{15}N-HSQC NMR spectrum was first recorded using 250 μM of ^{15}N-labelled $A3B_{CTD}$ in the 50 mM citrate-phosphate pH 5.5 buffer to entice interactions. This was followed by obtaining ^{15}N-HSQC NMR spectra titrated with differing ratios of protein to oligonucleotide (1:1, 1:2, 1:5, 1:10, 1:20).

2.5 Evaluation of the A3-catalysed reaction

2.5.1 *In vitro* DNA deamination gel-based assay

A previously reported *in vitro* DNA deamination in-gel based assay [240] was initially used to assess the deamination activity of the purified A3 proteins. Initially, the fluorescently-tagged oligonucleotide (TC-3' 6-FAM-oligo, see Table 2.13) and the purified proteins were buffer matched by diluting to required concentrations using a 1X HEPES pH 7.4 buffer (10 mM HEPES-KOH, 50 mM NaCl at pH 7.4). In a PCR strip, reactions were setup using 5 µL of a 1.6 µM TC-3' 6-FAM oligo mixed with 5 µL of 10 µM of the purified protein (final concentrations of 800 nM DNA to 5 µM protein). The reaction mixtures were then subjected to temperature cycling in a thermocycler (Eppendorf) as follows: first, the mixtures were incubated for an hour at 37 °C, followed by incubation at 95 °C for 3 minutes (to denature the protein and prevent further substrate deamination), then cooled to 12 °C. The reaction mixtures were then treated with 1 µL of 120 nM uracil-DNA glycosylase (UDG, which cleaves the uracil bases in the oligonucleotide) (New England Biolabs, Ipswich, MA USA) and incubated for 10 minutes at 37 °C. Following UDG treatment, 1.2 µL of a 1 M NaOH was added to the mixture and incubated for 5 minutes at 98 °C to stop the reaction and cleave DNA with an abasic site. To separate and resolve small fragments of the ssDNA reaction products a 15% denaturing PAGE gel was used.

A TBE-Urea PAGE gel was prepared by initially mixing 5.47 g of urea with 2.6 mL of a 5X TBE buffer (445 mM Tris at pH 7.6, 445 mM boric acid, and 10 mM EDTA) and dissolving at 50 °C. To this mixture 4.88 mL of a 40% acrylamide (for 15% gel), 6.5 µL of TEMED, and 91 µL of 10% ammonium persulfate were added. The resulting gel mixture (13 mL) was cast in a Mini-Protean II Electrophoresis Cell System (Bio-Rad) gel cassette at a thickness of 1.5 mm. Once the gel was polymerised, it was placed in the gel tank with 1X TBE buffer and run without samples for 20 minutes at 100 V to equilibrate the gel. Samples were mixed with 2X formamide buffer (80% formamide, 1X TBE, bromophenol blue and xylene cyanol) at a 1:1 ratio and 3 µL was loaded onto the gel. Electrophoresis was conducted at 200 V for 40 minutes (or until the bromophenol blue front line is near the bottom and xylene cyanol is halfway down the gel). The gel was washed in H_2O for 15 minutes before being scanned using fluorescence on a Typhoon FLA 7000 imager (GE Healthcare Life Sciences, Pittsburgh, PA USA).

2.5.2 1D ^{1}H-NMR-based activity assay

Kinetic characterisation of active $A3B_{CTD}$ variants ($A3B_{CTD}$-QM-ΔL3-AL1swap and $A3B_{CTD}$-DM) on their preferred cytosine in ssDNA sequences (Table 2.13Table 2.14) was quantitatively evaluated using a previously established 1D ^{1}H-NMR-based assay [79, 94]. This assay utilises the naturally abundant proton (^{1}H) nuclei within ssDNA oligonucleotides and monitors the real-time deamination of dC (substrate) to dU (product). Experiments were conducted on a 700-MHz Bruker NMR spectrometer equipped with a 1.7-mm cryoprobe at 298 K. ^{1}H NMR spectra of oligonucleotide standards containing dU were first recorded to determine where the H-5 proton doublet signal of the dU product appears. Initially, a series of ^{1}H NMR spectra of the oligonucleotide substrate was recorded at concentrations ranging from 50 μM to 750 μM with either 50 nM of $A3B_{CTD}$-QM-ΔL3-AL1swap or 2 μM of $A3B_{CTD}$-DM enzyme ([E] initial enzyme concentration), in a 50 mM citrate-phosphate buffer (50 mM citrate-phosphate, 200 mM NaCl, 2 mM β-ME, 200 μM DSS, pH 5.5 for $A3B_{CTD}$-QM-ΔL3-AL1swap, or pH 7.5 for $A3B_{CTD}$-DM) with 10% deuterium oxide (D_2O). A doublet of doublets occurring between 2.57 to 2.39 ppm originating from citrate in the buffer was used as an internal standard to determine the concentration of the substrate converted during the reaction. To determine the rate of deamination by the A3 enzyme, the H-5 proton doublet signal of the dC/dU was integrated then converted to a substrate concentration (termed [S]), which was plotted against the time of the reaction (sec) to derive the initial speed (also termed rate) of the reaction). From this $1/V_0$ was plotted against 1/[S] producing the Lineweaver-Burk plot, from which one can determine the Michaelis-Menten parameters. The Michaelis-Menten equation is described in line 1 of Equation 2.1, where V_{max} is described as the maximum rate when an enzyme is saturated with substrate, the Michaelis constant, K_m, describes the substrate's apparent binding affinity for the enzyme where [S] is at half of V_{max}. The catalytic rate constant, k_{cat}, describes the maximum number of substrate converted to product per enzyme per second (derived from line 2 in Equation 2.1, assuming V_{max} is reached independent of [S] at high [S]). To conveniently derive these parameters the Michaelis-Menten plot is converted to a Lineweaver-Burk plot by taking the inverse V_0 ($1/V_0$) against the inverse [S] (1/[S]) as described by Equation 2.1 line 3. The Lineweaver-Burk plot was fitted with a linear regression equation to simplify calculations of the Michaelis-Menten parameters (K_m and k_{cat}).

These parameters were calculated using Equation 2.2. First, the Lineweaver-Burk equation (line 1, Equation 2.2) was substituted into a linear regression equation (line 2, Equation 2.2) as shown in line 3 (Equation 2.2). Using this equation, *y* represents the dependent variable ($1/V_0$) and *x* was assigned as the independent variable (1/[S]). From the fitted line we can derive the value of the constant number *a* and *b*, where *a* is the *y*-intercept when $x = 0$, while *b* is equivalent to the slope of the line. Substituting values of *a* and *b* and rearranging each equation, k_{cat} (line 4, Equation 2.2) and K_m parameters are derived. Uncertainties of K_m and k_{cat} were calculated using the error-propagation method based on uncertainties in *a* and *b* from linear regression obtained using LINEST function in Excel described in the appendix (Equation 8.1).

$$\mathrm{y} = V_o = V_{max} \times \frac{[S]}{K_m + [S]} \quad (1)$$

$$V_{max} = k_{cat}[E] \quad (2)$$

$$\mathrm{y} = V_o = k_{cat}[E] \times \frac{[S]}{K_m + [S]} \quad (3)$$

$$\mathrm{y} = \frac{1}{V_o} = \frac{K_m + [S]}{k_{cat}[E][S]} \quad (4)$$

$$\mathrm{y} = \frac{1}{V_o} = \frac{K_m}{k_{cat}[E]} \times \frac{1}{[S]} + \left(\frac{1}{k_{cat}[E]}\right) \quad (5)$$

Equation 2.1 Michaelis-Menten and Lineweaver-Burk equations.

$$\mathrm{y} = \frac{1}{V_o} = \frac{K_m}{k_{cat}[E]} \times \frac{1}{[S]} + \left(\frac{1}{k_{cat}[E]}\right) \quad (1)$$

$$\mathrm{y} = ax + b \quad (2)$$

$$a = \frac{K_m}{k_{cat}[E]} \qquad b = \frac{1}{k_{cat}[E]} \quad (3)$$

$$k_{cat} = \frac{1}{b \times [E]} \quad (4)$$

$$K_m = \frac{a}{b} = \frac{K_m}{k_{cat}[E]} \div \frac{1}{k_{cat}[E]} \quad (5)$$

Equation 2.2 Calculation of K_m and k_{cat} from the Lineweaver-Burk equation using the linear regression equation.

2.5.3 1D ^{1}H-NMR-based inhibitor assay

NMR-based inhibition assays were carried out to determine the deaminase activity of A3 variants in the presence of the chemically modified ssDNA (Table 2.14) to assess the inhibitory potential. Similar to the NMR activity assay described above in section 2.5.2, a series of ^{1}H NMR spectra was recorded using 350 µM substrate oligo [S] along with varying concentrations of chemically modified ssDNA (5 µM to 100 µM) with either 50 nM of $A3B_{CTD}$-QM-ΔL3-AL1swap or 2 µM of $A3B_{CTD}$-DM in the buffers mentioned earlier in section 2.5.2. Integration of the H-5 proton doublet signal of the dC or dU was converted to substrate concentration and plotted against the time of the reaction to obtain the initial rate of the reaction (V_0). From this the inverse rate of the reaction ($1/V_0$) was plotted against the inhibitor concentration (termed [I]) to produce the Dixon plot at a given [S], as shown in line 1 Equation 2.3, under the assumption that our inhibitors are competitive inhibitors, as then validated experimentally in Figure 4.11.

The Dixon plot was fitted with a linear regression equation (terms described in section 2.5.2) to calculate the inhibition constant (K_i). The Dixon plot equation (line 1, Equation 2.3) was substituted into a linear regression equation (line 2, Equation 2.3) as shown in line 3 (Equation 2.3). Values of a (y-intercept), b (slope of line), and values of the Michaelis parameters obtained earlier (K_m and $k_{cat,}$ Equation 2.2) were substituted and the equation was solved. Uncertainty of K_i was calculated using error-propagation method from errors of least-square fit to the measured data as described in Equation 8.2 in appendix.

$$y = \frac{1}{V_o} = \frac{K_m}{K_i \times V_{max} \times [S]} \times [I] + \frac{K_m + [S]}{V_{max} \times [S]} \quad (1)$$

$$y = ax + b \quad (2)$$

$$a = \frac{K_m}{K_i \times V_{max} \times [S]} \qquad b = \frac{K_m + [S]}{V_{max} \times [S]} \quad (3)$$

$$V_{max} = k_{cat}[E] \quad (4)$$

$$K_i = \frac{K_m}{a \times k_{cat} \times [E] \times [S]} \quad (5)$$

$$[E] = \frac{K_m + [S]}{b \times k_{cat} \times [S]} \quad (6)$$

Equation 2.3 Calculation of K_i from the Dixon plot using the linear regression equation.

K_m and k_{cat} values are derived from Equation 2.2.

2.6 Structural studies

2.6.1 Analysis of the proteins structure using nuclear magnetic resonance (NMR) spectroscopy

To provide insight into a protein's structure, dynamics, and binding interactions in solution, multidimensional (2D and 3D) NMR experiments utilise radiofrequencies to detect the nuclei of proteins. Multidimensional experiments rely on the transfer of magnetisation energy between nuclei, providing specific connections between nuclei. To enable these experiments, proteins are artificially enriched with isotopes containing nuclei with an odd number of protons and/ neutrons which leads to a nuclear magnetic moment in the magnetic field (refer to preparation of isotopically labelled protein section 2.2.6) allowing assignment of signals in NMR spectra to atoms in the protein.

Several multidimensional NMR experiments listed in Table 8.1 (in the appendix) were conducted on $A3B_{CTD}$. Experiments were setup using 250 - 350 µM of labelled $A3B_{CTD}$ (^{15}N-labelled for 2D experiments or ^{15}N, ^{13}C-double-labelled for 3D experiments) in a volume of 300 µL high salt NMR buffer (50 mM Tris-HCl at pH 7.4, supplemented with 0.5 M NaCl, and 2 mM β-ME) along with 10% D_2O. To improve the signal to noise ratio, samples were measured in 5 mm high-throughput NMR tubes (Wilmad®). The sample was degassed for 30 minutes before being placed in a salt-tolerant susceptibility matched slot tube (Shigemi Inc.) and measured using a 700 MHz spectrometer (Bruker Avance) equipped with a 1.7-mm cryoprobe at 298 K.

2.6.2 Small-Angle X-ray Scattering (SAXS)

In the last few years, structures of the catalytically inactive single domain A3 bound to a ssDNA substrate were reported [129, 161, 239]. However, the structure of two-domain A3 in complex with ssDNA remains unclear due to the limitations of crystallography. Therefore, small-angle X-ray scattering (SAXS) was used to provide structural information on the protein and its complexes in solution. Low-resolution models of full-length A3-ssDNA obtained by SAXS can be combined with A3 domain structures to understand the arrangement of domains upon the binding DNA. Due to difficulties in producing full-length A3B, the closely related full-length A3G (provided by Linda Chelico from the University of Saskatchewan) was utilised.

Measurements were conducted at the Australian Synchrotron on the SAXS/WAXS beamline equipped with a Pilatus-2 1M detector. SAXS data were obtained at 25℃, using

a camera length of 1.6 m and frames were taken at one second intervals. Samples of A3 proteins (0.5 – 10 mg/mL) with and without ssDNA at varying ratios were loaded in a 96 well plate (50 µL) and degassed for 30 minutes. Scattering data were obtained using co-flow SEC-SAXS mode. Samples were injected onto a pre-equilibrated SEC column (Superdex® 200 Increase 5/150) then run at a flow rate of 0.2 mL/min while simultaneously SAXS measurements were acquired. All buffers listed in Table 2.15 contained 5-10% glycerol to limit radiation damage of samples. SEC UV elution profile recorded at various wavelengths (A_{280}, A_{260}, A_{220}, and A_{480}) over time (seconds), while SAXS data was collected per frame, but lagged 1.05 seconds/frame between the SEC elution profile.

Table 2.15 Buffers used in SAXS studies

Buffer name	Buffer composition
Kinetic pH 5.5 buffer (for A3B)	50 mM citrate-phosphate pH 5.5, 200 mM NaCl, 2 mM β-ME, 200 µM DSS, 10% glycerol
High salt pH 7.4 buffer (for A3B)	50 mM Tris-HCl pH 7.4, 0.5 M NaCl, and 2 mM β-ME, 10% glycerol
A3G pH 6 buffer	50 mM phosphate pH 6.0, 200 mM NaCl, 2mM β-ME, 5% glycerol, 200 µM Na_2EDTA

Raw SAXS data frames were first reduced by intensity normalisation, background subtraction, and scattering vector calibration using SCATTERBRAIN 2.82 program (http://www.synchrotron.org.au/aussyncbeamlines/saxswaxs/software-saxswaxs). A SAXS profile plot was then derived by plotting the normalised integrated intensity of the scattering signal against the frame number using CHROMIX (ATSAS 2.8.3 suite [75]). The frames of interest are averaged and are subtracted from the buffer frames using CHROMIX. Then the processed scattering data are transformed into a 1D-scattering curve using the PRIMUSQT program from ATSAS 2.8.3 suite [75], where log of scattering intensity (log(I)) was plotted against the scattering vector (q) ($q = 4\pi\sin(\theta/\lambda)$, 2θ is the scattering angle, and λ is the x-ray wavelength of 1.0332 Å). When data are partially resolved they underwent evaluation using singular value decomposition (SVD) and evolving factor analysis (EFA) (BioXTAS RAW [105] or US-SOMO HPLC-SAXS module [31-33]) to determine the number and boundaries of each component in the sample. Where

data were not resolved, they must undergo deconvolution using Gaussian decomposition methodologies through US-SOMO HPLC-SAXS module [31-33].

Characteristic parameters can be retrieved from the scattering pattern of the samples that describe the homogeneity, fold, size, and overall shape of the sample. Several important parameters derived include the radius of gyration (*R*g, Å), molecular weight (MW, kDa), excluded particle volume (Porod volume) (*V*, Å^3), maximum dimension ($D_{max,}$ Å), and the extrapolated intensity at zero scattering angle ($I(0)$, cm^{-1}) obtained by analysis of the scattering curve using several programs that are part of the ATSAS package (programs include PRIMUSQT, AUTORG, DATPOROD, and AUTOGNOM) [75]. Errors reported for parameters refers to a weighted least square fit of the intensity. The scattering curve can be compared to generated 1D-scattering curves calculated from high resolution structures using either CRYSOL [75] or the FoXS server [74, 226]. Envelope dummy atom models can additionally be generated using DAMMIF from the pair distribution function ($P(r)$) of the data (ATSAS 2.8.3 suite [75]), and superimposed with high resolution structures using PyMol [228]. SAXS analysis described in more detail in section 6.1.

3. Evaluation of the binding affinity of ssDNA to $A3B_{CTD}$ protein

3.1 Introduction

Some of the work reported in this chapter was presented in the conference listed below.

> Barzak, F. M., Harjes S., Filichev, V. V., Harjes, E. Stopping cancer evolution: towards understanding of APOBEC3B-DNA interactions. Poster presentation at the 9th Australian and New Zealand Magnetic Resonance Society (ANZMAG) conference 2015 held between the 29th of November - 3rd of December 2015 at the Copthorne hotel and resort, Paihia, Bay of Islands, New Zealand.

At the time of the research conducted in this chapter there were no available structures of A3 proteins in complex with ssDNA, only ligand-free structures [22, 36, 94, 103, 233]. Therefore, we initially aimed to structurally characterise the A3B-ssDNA complex to understand the molecular basis of interactions to provide structurally informed design of drugs that specifically target A3B. Reported difficulty in the expression and purification of the full-length recombinant two-domain A3B prompted the characterisation of the CTD of A3B [240], as this domain constitutes the deaminase activity [78, 283].

In the beginning, we focused on producing sufficient quantities of soluble $A3B_{CTD}$ proteins to explore the molecular basis of interactions and obtain the structure of the $A3B_{CTD}$-ssDNA complex. Catalytically inactive (on the experimental time scale) $A3B_{CTD}$ variants were selected to avoid deamination of ssDNA containing dC during binding studies. Substrate oligonucleotide sequences specific for A3B were designed (described in section 2.3.1) and evaluated for their binding affinity towards $A3B_{CTD}$. Initially, ssDNA sequences were screened for their binding capability to the purified $A3B_{CTD}$ protein variants. Then the chosen oligonucleotide was examined using biophysical experiments to understand the binding interactions between the $A3B_{CTD}$ protein and ssDNA. In an attempt to characterise the binding location of ssDNA onto the surface of the $A3B_{CTD}$, solution-state NMR was first used to assign the $A3B_{CTD}$ spectrum, then chemical shift perturbations were monitored upon titration with ssDNA to potentially identify the oligonucleotide's binding site. However, during this study the crystal structures of the $A3B_{CTD}$ and the closely related A3A in complex with ssDNA were published [239-240].

3.2 Protein expression and purification of A3B$_{CTD}$ variants: A3B$_{CTD}$-QM-ΔL3 and A3B$_{CTD}$-QM-ΔL3-E255A

A3B$_{CTD}$ variant constructs; A3B$_{CTD}$-QM-ΔL3 and A3B$_{CTD}$-QM-ΔL3-E255A, cloned into pET24a vector to obtain C-terminally His$_6$-tagged proteins (listed in Table 2.2), were used for this study [240]. These variants contained four substitution mutations of hydrophobic amino acids, situated away from the active-site, to hydrophilic residues (F200S, W228S, L230K, F308K), and the removal of the flexible loop 3 (Figure 3.1). The A3B$_{CTD}$-QM-ΔL3-E255A variant had an additional modification where the active-site glutamic acid (E255A) was replaced with an alanine, making the protein catalytically inactive (Figure 3.1).

```
A3BCTD                --------------PDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTWVLMDQHMGFLCN
A3BCTD-QM-ΔL3         ------EILRYLMDPDTFTSNFNNDPLVLRRRQTYLCYEVERLDNGTSVKMDQHMGFLCN
A3BCTD-QM-ΔL3-E255A   ------EILRYLMDPDTFTSNFNNDPLVLRRRQTYLCYEVERLDNGTSVKMDQHMGFLCN
                                          *                              * *

A3BCTD                EAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQEN
A3BCTD-QM-ΔL3         E--------SGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQEN
A3BCTD-QM-ΔL3-E255A   E--------SGRHAALRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQEN
                       *********    *

A3BCTD                THVRLRIFAARIYDY-DPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPW
A3BCTD-QM-ΔL3         THVRLRIKAARIYDY-DPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPW
A3BCTD-QM-ΔL3-E255A   THVRLRIKAARIYDY-DPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPW
                             *

A3BCTD                DGLEEHSQALSGRLRAILQNQGN
A3BCTD-QM-ΔL3         DGLEEHSQALSGRLRAILQ----
A3BCTD-QM-ΔL3-E255A   DGLEEHSQALSGRLRAILQ----
```

Figure 3.1 A3B$_{CTD}$ variants amino acid sequence used for binding studies.

A3B$_{CTD}$ residues 187 to 388. A3B$_{CTD}$-QM-ΔL3 (Quadra Mutant, delta loop 3): F200S, W228S, L230K, and F308K, loop 3 deletion. A3B$_{CTD}$-QM-ΔL3-E255A: F200S, W228S, L230K, and F308K, loop 3 deletion, active-site glutamate mutated to alanine (E255A). Constructs kindly provided by Assoc. Prof. Hideki Aihara and Prof. Reuben Harris, University of Minnesota [240].

3.2.1 Transformation of A3B$_{CTD}$ constructs (A3B$_{CTD}$-QM-ΔL3 and A3B$_{CTD}$-QM-ΔL3-E255A)

Expression of the recombinant A3B$_{CTD}$ variants, A3B$_{CTD}$-QM-ΔL3 and A3B$_{CTD}$-QM-ΔL3-E255A, began by transforming constructs into chemically competent *E. coli* cells (as described in section 2.2.3). The A3B$_{CTD}$ plasmid variants isolated from the transformed colonies were run on a 1% agarose gel as described in section 2.2.3.1, confirming that an intact plasmid with the expected recombinant plasmid size of approximately 6 kb had been obtained (pET24a recombinant plasmid ~ 5.4 kb, A3B$_{CTD}$ 187-387 amino acid sequence insert 0.6 kb) as shown in Figure 3.2. The isolated plasmids were then sequenced using the method described in section 2.2.3.2, confirming that the

correct amino acid sequence of the $A3B_{CTD}$ variant was present in the plasmids that were isolated. The isolated plasmids were then transformed into a protein-expressing BL21 (DE3) *E. coli* strain (Lucigen) as described in section 2.2.4.

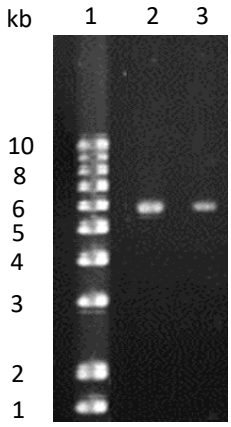


Figure 3.2 Agarose gel (1%) of the isolated $A3B_{CTD}$ plasmid variants.

Lane 1: 1kb plus DNA standard ladder, lane 2: pET24a-$A3B_{CTD}$-QM-ΔL3, lane 3: pET24a-$A3B_{CTD}$-QM-ΔL3-E255A. pET24a plasmid size is 5.4 kb and $A3B_{CTD}$ insert is 0.6 kb.

3.2.2 Expression of the $A3B_{CTD}$ constructs ($A3B_{CTD}$-QM-ΔL3 and $A3B_{CTD}$-QM-ΔL3-E255A)

Initially, small-scale expression trials of the $A3B_{CTD}$ variants, $A3B_{CTD}$-QM-ΔL3 and $A3B_{CTD}$-QM-ΔL3-E255A, were conducted to assess the level of soluble protein production. Note that expression and purification of both variants were carried out but for optimisation was conducted on the $A3B_{CTD}$-QM-ΔL3 variant. The $A3B_{CTD}$ constructs were mainly expressed and purified as previously reported [239-240] with minor optimisations detailed here. Briefly, the recombinantly transformed $A3B_{CTD}$ culture was cultivated in LB medium at 37 °C until OD_{600} ~ 0.6 (mid-log phase) was established. The culture was then supplemented with zinc chloride to a final concentration of 100 µM and protein expression induced by IPTG at a final concentration of 0.5 mM, followed by incubation at 18 °C overnight as described in section 2.2.4. Examination of the overnight cultures on a tricine SDS-PAGE gel stained with Coomassie blue (see materials and methods 2.2.7.2), showed that upon IPTG induction an overexpressed protein band appears estimated to be around 22 kDa in size (arrow in Figure 3.3). The size of the overexpressed protein was consistent with the estimated molecular weight of the $A3B_{CTD}$ protein of 23 kDa, and since the protein was only expressed upon activation of the T7

promoter system by IPTG, it was presumed that this $A3B_{CTD}$ protein variant had been obtained.

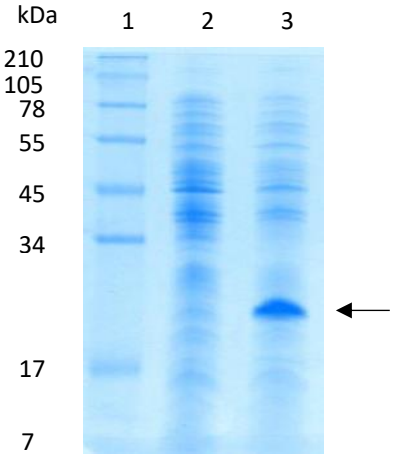


Figure 3.3 Tricine SDS-PAGE gel (16%) of $A3B_{CTD}$-QM-ΔL3 protein expression trial.

Lane 1: SeeBlue Plus2 Pre-stained protein standard ladder, lane 2: uninduced cell culture of $A3B_{CTD}$-QM-ΔL3, lane 3: induced cell culture of $A3B_{CTD}$-QM-ΔL3 with final concentration of 0.5 mM IPTG.

To confirm the identity of the overexpressed protein attained upon IPTG induction (Figure 3.3), mass spectrometry was employed. First, the induced sample was run on a tricine SDS-PAGE gel, which was stained with colloidal Coomassie stain and in-gel tryptic digest was carried out on the protein band of interest, as outlined in section 2.2.7.3. Samples were then run on ESI-QUAD-TOF mass spectrometer and analysed by Mr. Trevor Loo (2.2.7.3). Analysis reports with peptides scoring greater than 39 of individual ions indicates identity or extensive homology ($p<0.05$). The protein band of interest had four high quality unique peptide hits associated with the A3B C-terminal region highlighted in Figure 3.4, with ion scores greater than 39, concluding that the protein band expressed was in fact $A3B_{CTD}$.

190 200 210 220 230 240 250 260

EILR YLMDPDTFT**S** NFNNDPLVLR RRQTYLCYEV ERLDNGT**S**V**K** MDQHMGFLCN E--------**S** GRHAELRFLD

270 280 290 300 310 320 330

LVPSLQLDPA QIYRVTWFIS WSPCFSWGCA GEVRAFLQEN THVRLRI**K**AA RIYDYDPLYK EALQMLRDAG

340 350 360 380

AQVSIMTYDE FEYCWDTFVY RQGCPFQPWD GLEEHSQALS GRLRAILQ

Figure 3.4 In-gel tryptic digest mass spectrometry analysis of the $A3B_{CTD}$-QM-ΔL3 protein band.

$A3B_{CTD}$-QM-ΔL3 sequence with four unique peptides (29% coverage) identified by mass spectrometry. Peptides highlighted with four different colours.

Additional expression trials were conducted to determine if the expressed $A3B_{CTD}$-QM-ΔL3 protein construct was soluble. The $A3B_{CTD}$ variant was expressed as described earlier (see above), followed by harvesting the cell culture by centrifugation. The cells were lysed using a high salt buffer containing 50 mM Tris-HCl pH 7.4, 500 mM NaCl, and 2 mM β-ME with 0.5 mg/mL lysozyme, sonicated, then the whole cell lysate was pelleted. Assessment of the fractions on tricine SDS PAGE gel revealed that the protein was partly soluble; however, a considerable proportion of the protein was insoluble remaining in the pellet fraction (see buffer A, Figure 3.5).

In an attempt to enhance the solubility of the $A3B_{CTD}$ protein, several protein induction conditions, such as induction temperature (15 – 25 °C), IPTG concentration (0.2 mM - 1 mM), and Zn^{2+} concentration (20 - 100 µM), were trialled. However, no notable improvements were obtained. Next, various lysis buffer compositions were trialled to determine if they can improve the release of soluble proteins. Upon changes in osmotic pressure caused by a decrease in the salt concentration of the lysis buffer from 500 mM to 150 mM NaCl, the cells lysed more efficiently leading to a significant improvement in the solubility of the protein, seen by the major increase in the band density of the supernatant fraction buffer B in comparison to buffer A (Figure 3.5).

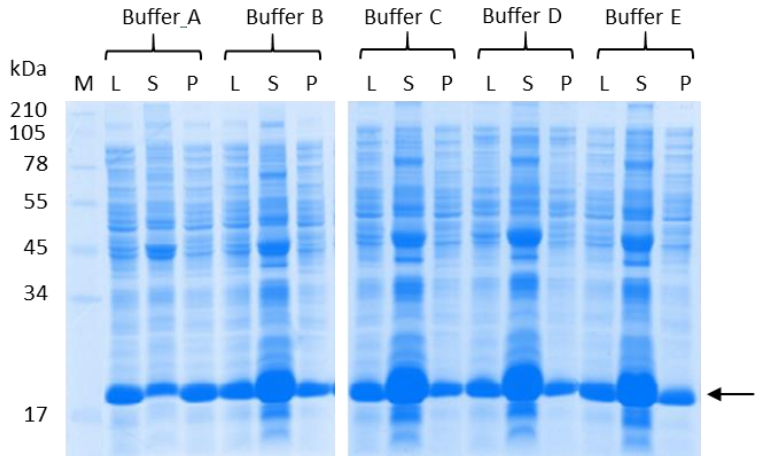


Figure 3.5 Tricine SDS-PAGE (16%) gel of a lysis buffer screen for the extraction of $A3B_{CTD}$-QM-ΔL3.

$A3B_{CTD}$-QM-ΔL3 cell culture treated with lysis buffers A-E. Loaded samples: M: SeeBlue Plus2 Pre-stained protein standard ladder, L: cell lysate at a 1:2 sample to dye ratio, S: supernatant at a 1:1 sample to dye ratio, P: pellet at a 1:1 sample to dye ratio. Buffer A: 50 mM Tris-HCl pH 7.4, 500 mM NaCl, and 2 mM β-ME with 0.5 mg/mL lysozyme, Buffer B: 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, and 2 mM β-ME with 0.5 mg/mL lysozyme, Buffer C: 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, and 2 mM β-ME, 0.1 % TWEEN with 0.5 mg/mL lysozyme, Buffer D: 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, and 2 mM β-ME, 1 mM EDTA with 0.5 mg/mL lysozyme, Buffer E: 50 mM HEPES, pH 7.4, 150 mM NaCl, and 2 mM β-ME, with 0.5 mg/mL lysozyme.

Trials to further enhance the efficacy of the lysis buffer by either adding a detergent (0.1% TWEEN), a chelating agent (1 mM Na_2-EDTA) or switching the buffering agent (50 mM HEPES pH 7.5 buffer) were conducted. However, this resulted in no further improvements in the proteins' solubility as observed by the band eluting in the supernatant of Buffers C-E in Figure 3.5. Therefore, lysis buffer B with the lower salt concentration was selected for future protein extractions.

3.2.3 Purification of the A3B$_{CTD}$ constructs (A3B$_{CTD}$-QM-ΔL3 and A3B$_{CTD}$-QM-ΔL3-E255A)

Next, a large scale protein expression of the A3B$_{CTD}$ variants (A3B$_{CTD}$-QM-ΔL3 and A3B$_{CTD}$-QM-ΔL3-E255A) was carried out, induced, harvested, and the clarified soluble cell lysate was obtained for protein purification (as in section 2.2.4). Purification of the histidine-tagged A3B$_{CTD}$ proteins (A3B$_{CTD}$-QM-ΔL3 and A3B$_{CTD}$- QM-ΔL3-E255A) was conducted as described in section 2.2.5. Originally, it was vital to carry out various iterations to develop optimum conditions to purify the A3B$_{CTD}$ proteins. The collected fractions were analysed by tricine SDS-PAGE revealing that most of the contaminating proteins were removed in the initial wash stages observed in lanes 6-8 of Figure 3.6. The A3B$_{CTD}$ protein variant specifically bound to the Ni^{2+}-NTA matrix and eluted from the column upon washing with buffer containing higher imidazole concentrations (200-300 mM), as indicated by the presence of a 22 kDa protein band corresponding to the size of the protein of interest in lanes 9-10 of Figure 3.6.

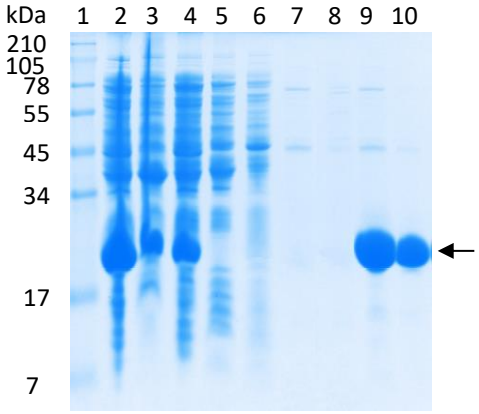


Figure 3.6 Tricine SDS-PAGE (16%) gel of IMAC protein purification of the A3B$_{CTD}$-QM-ΔL3 variant.

Lane 1: SeeBlue Plus2 Pre-stained protein standard ladder, lane 2: cell lysate, lane 3: pellet, lane 4: supernatant, lane 5: supernatant flow-through off column, lanes 6-8: buffer wash fractions *, lanes 9-10: eluted fractions $^{+}$. *Wash buffer: 50 mM Tris-HCl pH 7.4, 0.5 M NaCl, 2 mM β-ME, with 10 mM, 50 mM, or 100 mM imidazole. $^{+}$Elution buffer: 50 mM Tris-HCl pH 7.4, 0.5 M NaCl, 2 mM β-ME, with 200 mM, or 300 mM imidazole.

Following IMAC, the fractions collected that contained the A3B$_{CTD}$ were pooled together, then dialysed in high salt pH 7.4 buffer to remove residual imidazole. The dialysed sample was further purified using SEC-FPLC to remove any contaminating proteins as described in section 2.2.5. SEC resulted in the appearance of a single peak eluting after 90 mL of buffer flow through as seen in the chromatogram in Figure 3.7. Based on the appearance of a single eluting peak during SEC-FPLC purification and its eluting retention volume it was confirmed that these A3B$_{CTD}$ variants (A3B$_{CTD}$-QM-ΔL3 and A3B$_{CTD}$-QM-ΔL3-E255A) behaved as monomers. Fractions from the eluted peak were examined by tricine SDS-PAGE (2.2.7.2), confirming the presence of a clean A3B$_{CTD}$ band running at 22 kDa. The A3B$_{CTD}$ variants discussed in this chapter (A3B$_{CTD}$-QM-ΔL3, and A3B$_{CTD}$-QM-ΔL3-E255A) were successfully purified in high quantities (mg/mL) for use in biophysical studies.

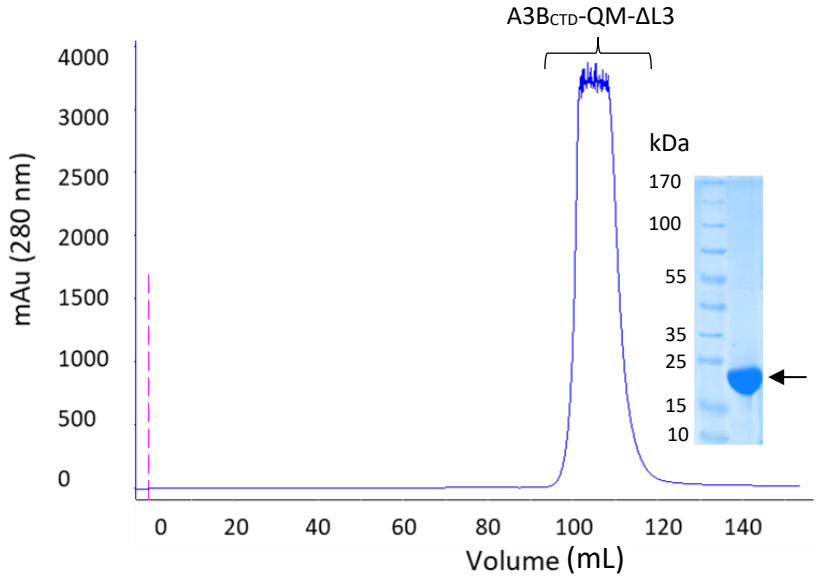


Figure 3.7 Size exclusion chromatography showing the protein purification of A3B$_{CTD}$-QM-ΔL3. A3B$_{CTD}$-QM-ΔL3 purified through a Superdex 75 10/300 GL column at a flow rate of 0.5 mL/min. Insert shows fraction of the SEC purified A3Bctd-QM-ΔL3 protein on a Tricine SDS-PAGE (16%) gel.

3.3 Examination of the catalytic activity of the purified A3B$_{CTD}$ constructs (A3B$_{CTD}$-QM-ΔL3 and A3B$_{CTD}$-QM-ΔL3-E255A) using an NMR-based activity assay

A fluorescence-based *in vitro* deaminase activity assay originally reported that the A3B$_{CTD}$-QM-ΔL3 variant was catalytically active [240]. Using this method described in section 2.5.2, our purified A3B$_{CTD}$-QM-ΔL3 variant was assessed, confirming these observations (see appendix Figure 8.2). However, to quantitively characterise the

deamination rate of the $A3B_{CTD}$ enzymes an established NMR-based activity assay was used [79, 94] (described in section 2.5.1). This assay monitors the real-time deamination of dC (substrate) to dU (product) in ssDNA by exploiting the naturally abundant proton (^1H) isotope within ssDNA using one-dimensional proton-NMR spectroscopy (^1H-NMR). Kinetic studies can be conducted by observing the formation of the proton (^1H) product signal or disappearance of the ^1H substrate signal over time. The ^1H signal appears as a peak (or peaks) at a specific position on the δ scale (in ppm) defined as a chemical shift (Figure 3.8).

The A9C oligonucleotide (5'-AATTCAAAA) was selected for deamination assessment by $A3B_{CTD}$-QM-ΔL3, as it had the best binding capability to $A3B_{CTD}$ (discussed further in section 3.4.3). First, ^1H-NMR spectra of the substrate oligonucleotide (A9C-oligo, 5'-AATTCAAAA) and product oligonucleotide (A9U-oligo, 5'-AATTUAAAA) were recorded to look for suitable chemical shifts to distinguish between the substrate and the product. The H-5 proton doublet signal of dC appears at 5.84 ppm (J = 7.5 Hz), while the H-5 proton doublet of dU is present at 5.665 ppm (J = 7.65 Hz) (Figure 3.8A and B i) (Table 8.5). Those signals were used in previous studies [79, 94]. However, we noted that the observed H-5 chemical shift of dU overlapped with another proton signal. The H-5 proton doublet of dU appeared to arise within a doublet of doublet signal (between 5.70 – 5.65 ppm) corresponding to the chemical shift of H-1 of 2'-deoxyadenosine (dA) within the sequence. Therefore, an additional region of the spectrum (between 7.43 – 7.32 ppm) was utilised as an indicator of cytosine deamination (Figure 3.8A ii and B ii). In this region, a doublet signal corresponding to H-4 of 2'-deoxythymidine (dT) is present at 7.41 ppm (J = 7.72 Hz), as well as a single peak resultant of H-6 of dA at 7.345 ppm within the A9C-oligo sequence. Upon deamination the H-4 dT doublet disappears, while the H-6 dA peak shifts slightly to 7.345 ppm as indicated using the A9U-oligo (product) standard sequence (Figure 3.8A ii and B ii).

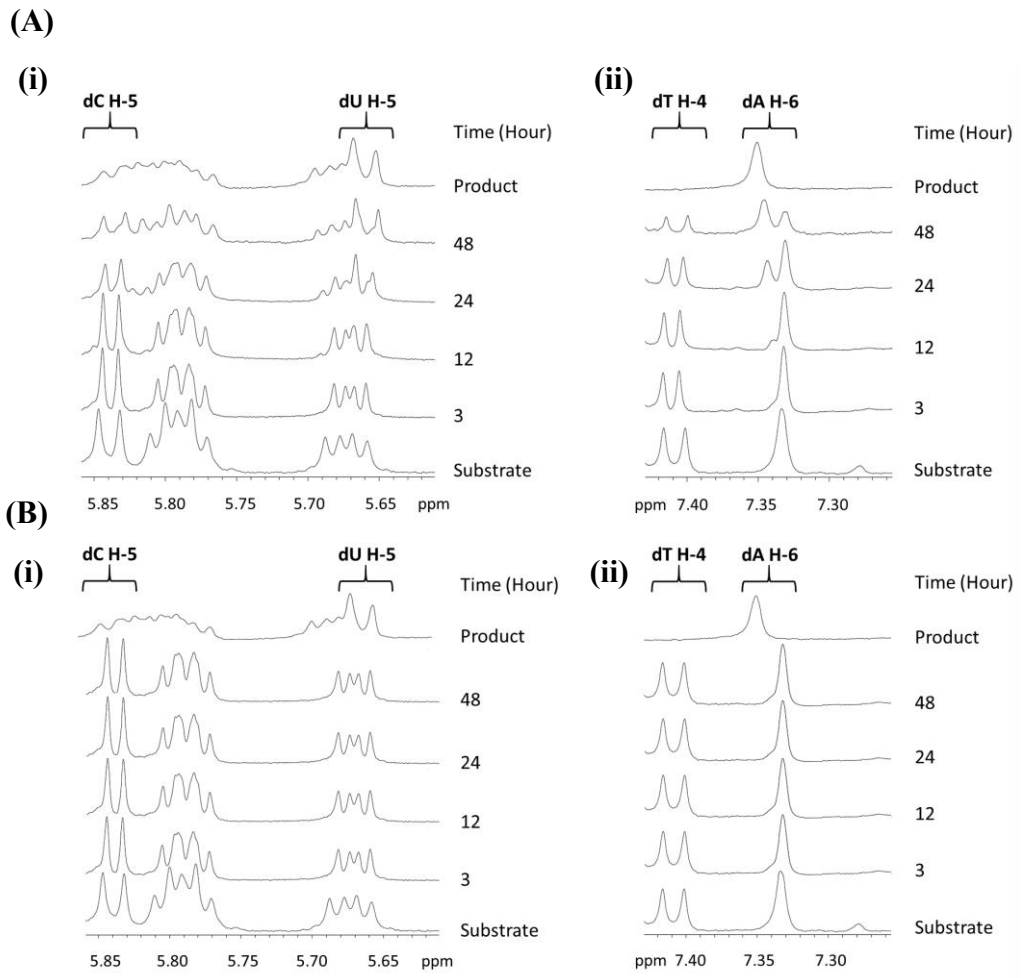


Figure 3.8 1D ^{1}H-NMR-based activity assay of the deamination of dC in ssDNA by A3B$_{CTD}$ variants over time.

(A) Assay using A3B$_{CTD}$-QM-ΔL3, (B) or A3B$_{CTD}$-QM-ΔL3-E255A. (i) ^{1}H-NMR spectrum 5.85-5.65 ppm, (ii) ^{1}H-NMR spectrum 7.40-7.30 ppm. Assay setup using 1 mM substrate A9C-oligo (5'-AATTCAAAA) along with 100 µM A3B$_{CTD}$ in a 50 mM citrate-phosphate, pH 7.5 buffer (50 mM citrate-phosphate buffer, 200 mM NaCl, 2 mM β-ME, 200 µM DSS, pH 7.5, with 10% D_2O) and monitored over 48 hours. Product A9U-oligo (5'-AATTUAAAA).

Next, the assay was set up using 1 mM of substrate oligonucleotide (A9C-oligo) incubated along with 100 µM of the purified A3B$_{CTD}$ enzyme in a 50 mM citrate-phosphate pH 7.5 buffer. Examination of the NMR spectrum after 12 hours of conducting the reaction showed that there was only a minor conversion of dC to dU, as indicated by the appearance of a rather small peak corresponding to H-6 dA (7.345 ppm) (Figure 3.8A ii). After incubation overnight (24 hours) only half of dC was converted to dU, as shown by a reduction in the H-5 dC doublet and the emergence of the H-5 dU doublet (Figure 3.8A i). In contrast to the aforementioned study [239], this recombinantly purified A3B$_{CTD}$-QM-ΔL3 protein was comparatively inactive as indicated by the slow deamination of the

oligonucleotide. On the other hand, the recombinantly purified $A3B_{CTD}$-QM-ΔL3-E255A protein led to no chemical shift changes (Figure 3.8B), validating that it is catalytically inactive. These $A3B_{CTD}$ variants ($A3B_{CTD}$-QM-ΔL3 and $A3B_{CTD}$-QM-ΔL3-E255A) are unsuitable for catalytic studies. However, the lack of cytosine deamination is useful for evaluating the binding behaviour of $A3B_{CTD}$ to ssDNA to prevent substrate conversion during the experiment.

3.4 Binding affinity of ligands to $A3B_{CTD}$ variants examined by protein thermal stability

The protein's thermal stability was examined using an established fluorescence-based thermal shift assay [93] to evaluate the biomolecular interactions between $A3B_{CTD}$ proteins and ligands (ssDNA in our case) as described in section 2.4.1. This assay relies on monitoring the fluorescent signal upon thermal denaturation of the protein. As the protein unfolds due to heat exposure, a fluorescent dye (SYPRO® orange dye, BioRad) binds non-specifically to the exposed hydrophobic regions of the protein, which leads to an increase in the fluorescent signal until the protein is completely denatured. The protein's melting temperature (T_m) is derived from the thermal denaturation curve where the protein transitions between folded and unfolded state, defined as the midpoint (Figure 3.9). An increase in T_m upon the presence of a ligand is an indication of the formation of a complex that stabilises the protein's conformational state.

To promote formation of an $A3B_{CTD}$-ssDNA complex several conditions were employed. First, the experiment was setup using 100 µM oligonucleotide to 20 µM protein [59]. Second, a mildly acidic buffer was utilised (50 mM citrate-phosphate buffer at pH 5.5, containing 200 mM NaCl, 2 mM β-ME, and 200 µM DSS) to favour formation of electrostatic interactions and hydrogen bonding between the ssDNA phosphate backbone and the amino acids in $A3B_{CTD}$ loop regions. Lastly, catalytically inactive or weakly inactive $A3B_{CTD}$ proteins ($A3B_{CTD}$-QM-ΔL3 and $A3B_{CTD}$-QM-ΔL3-E255A, assessed earlier in section 3.3) were used to prevent the catalysis of the ssDNA substrate during the experiment. Assays were setup as outlined in section 2.4.1. Thermal shift experiments (Figure 3.9 - Figure 3.12) showed that the presence of substrate ssDNA led to a small increase in the melting temperature (ΔT_m) of the $A3B_{CTD}$ (maximum $\Delta T_m \sim 2.0$ °C). In contrast, S. Harjes et al [93] reported that the catalytically inactive A3A enzyme (A3A-E72A) had a much stronger binding affinity towards ssDNA, as indicated by a larger ΔT_m

~ 4.3 °C using the same method. This implied that the A3B$_{CTD}$ binds ssDNA with a relatively weak affinity in comparison to A3A [93, 135]. From the structure, it was found that loop 3 interacts with the phosphate backbone of the ssDNA +1 nucleotide through a hydrogen bond to stabilise the oligonucleotide's binding conformation [93, 239]. Therefore, deletion of loop 3 in our A3B$_{CTD}$ variants may account for this observed weaker binding affinity. Moreover, in contrast to the ligand-free A3A and ligand-free A3G$_{CTD}$ structures, the ligand-free A3B$_{CTD}$ has a tightly closed active site formed through stacking interactions between the side chains in loops 1 and 7 that require comprehensive rearrangements to bind ssDNA, thereby leading to a weaker binding affinity [239-240].

3.4.1 Catalytic glutamic acid determines binding of ssDNA to A3

To investigate A3 enzymes bound to ssDNA, our catalytically inactive (under our conditions) A3B$_{CTD}$ variants were used to prevent complications of catalysis [129, 239]. Kouno et al, showed that the replacement of the catalytic glutamic acid with alanine causes an inactive A3 enzyme, but does not alter the shape of the A3 active-site. The crystal structure showed that in place of the glutamate side chain a water molecule was present, suggesting that the active-site would still be accessible to bind to the oligonucleotide [129]. Binding assays were performed using two A3B$_{CTD}$ variants, A3B$_{CTD}$-QM-ΔL3 and A3B$_{CTD}$-QM-ΔL3-E255A, that differed by a single point mutation (E255A) to assess the importance of the catalytic glutamate (E255) in binding ssDNA. The proteins' thermal stability was evaluated as described earlier (see section 3.4).

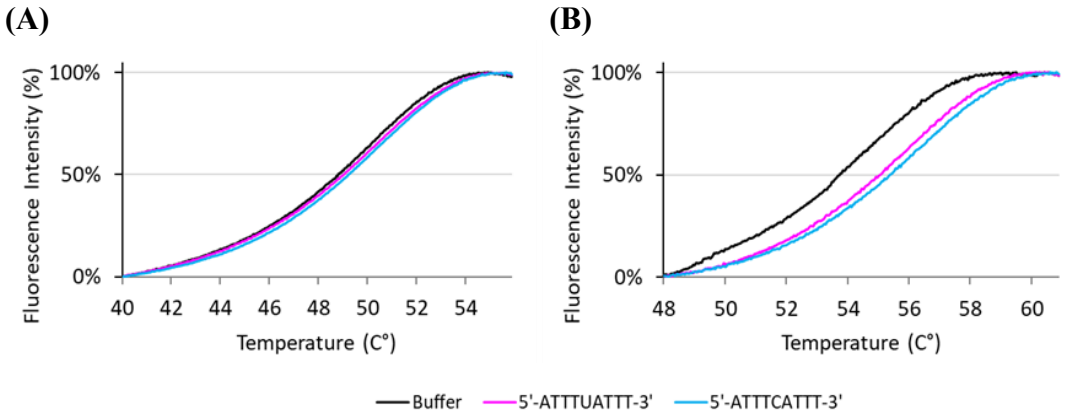


Figure 3.9 Thermal denaturation profiles of A3B$_{CTD}$ variants.

Thermal shift assay of (A) A3B$_{CTD}$-QM-ΔL3-E255A and (B) A3B$_{CTD}$-QM-ΔL3 in the presence of buffer (black), dU-containing ssDNA (5'-ATTTUATTT, magenta), dC-containing ssDNA (5'-ATTTCATTT, blue) run in a 50 mM citrate-phosphate, pH 5.5 buffer containing 10X SYPRO® orange dye (BioRad). Data were normalised to percentage of fluorescence intensity plotted against temperature, to calculate the T_m midpoint.

At first glance, the presence of ssDNA leads to an observable increase in the melting temperature (T_m) of the $A3B_{CTD}$ variants, but more prominently for the $A3B_{CTD}$-QM-ΔL3 derivative (Figure 3.9). More specifically, the substrate oligonucleotide containing dC had a slightly higher change in the melting temperature (ΔT_m) than the product oligonucleotide containing dU, as a direct consequence of a lower binding affinity towards the product of deamination [23, 93, 245]. On closer inspection, we showed that the substrate binds considerably tighter to the $A3B_{CTD}$-QM-ΔL3 (ΔT_m = 1.73 ± 0.21 °C, mean ± SD) (Figure 3.9B) than to the $A3B_{CTD}$-QM-ΔL3-E255A (ΔT_m = 0.30 ± 0.05 °C, mean ± SD) (Figure 3.9A). Moreover, the $A3B_{CTD}$-QM-ΔL3-E255A T_m values (48.9 - 49.1 °C) were comparatively lower than the $A3B_{CTD}$-QM-ΔL3 derivative (53.9 - 55.6 °C) with the intact catalytic glutamate in the active site. This indicates that the catalytic glutamate (E255) is essential in forming specific interactions with the oligonucleotide. As well as causing a decrease in thermal stability of the protein, the absence of the catalytic glutamate causes the loss of specific hydrogen bond interactions that lead to the lack of deamination of the target dC, consistent with the cytosine deamination model where glutamic acid initiates deamination by protonating the N3 of cytosine [100, 120, 188]. These results highlight the significance of having an intact catalytic glutamate in binding to ssDNA.

3.4.2 Binding of dZ-modified oligonucleotides to inactive $A3B_{CTD}$ variants

It was assumed that cytosine analogues would mimic the transition state of cytosine deamination and therefore bind to A3 enzymes with a stronger affinity than the substrate (dC) or product (dU), by converting the activation energy required for catalysis into static binding energy. Therefore, a chemically modified oligonucleotide containing 2'-deoxyzebularine (dZ) (see Figure 3.10A) in place of the target cytosine (5'-ATTTdZATTT, see 2.3.2) was assessed to determine its binding capability to our inactive $A3B_{CTD}$ variants, using the thermal shift assay (as described in 3.4.1). This assay showed that the modified oligo containing dZ bound better to the $A3B_{CTD}$-QM-ΔL3 protein than did the substrate oligo containing dC as reflected by an increase in the stability of the protein (ΔT_m = 0.72 ± 0.10 °C) (Figure 3.11). This result is consistent with cytosine analouges mimicking the transition state or an intermediate of the cytosine deamination reaction. The lack of the catalytic glutamate (E255) resulted in a poor binding affinity of the dZ-containing oligo with ΔT_m values similar to those in the presence of dC and dU-

containing oligos. Therefore, the A3B$_{CTD}$-QM-ΔL3 variant was selected for all further binding studies.

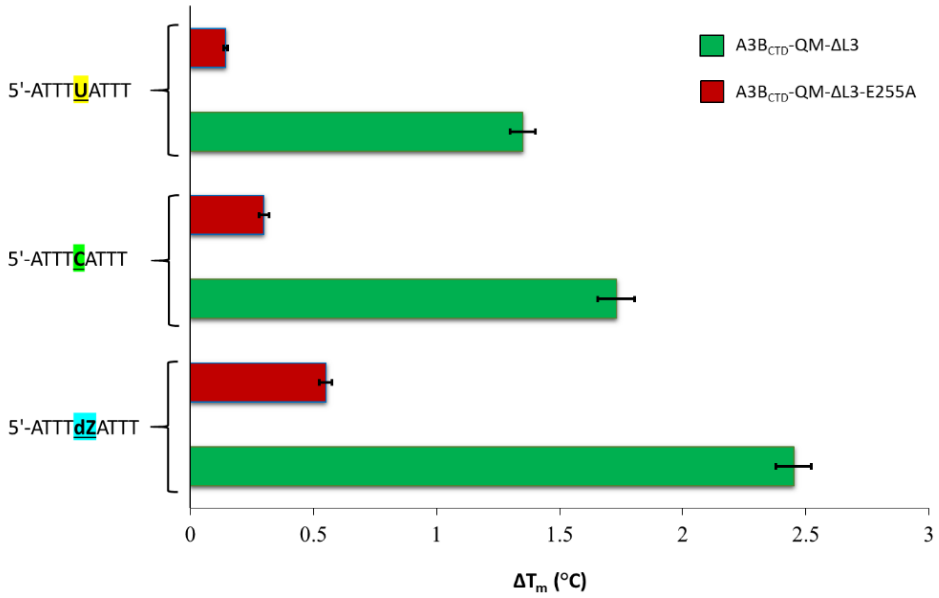


Figure 3.11 Binding capability of ssDNA to A3B$_{CTD}$ variants using thermal shift assay.
A3B$_{CTD}$-QM-ΔL3 (green) and A3B$_{CTD}$-QM-ΔL3-E255A (red) in the presence of a dU-containing oligo (5'-ATTTUATTT), dC-containing oligo (5'-ATTTCATTT), or dZ-containing oligo (5'-ATTTdZATTT). The assay was conducted using 20 μM protein with 100 μM oligo in the 50 mM citrate-phosphate pH 5.5 buffer (50 mM citrate-phosphate, 200 mM NaCl, 2 mM β-ME, and 200 μM DSS, pH 5.5). Confidence intervals (95%) are shown as error bars. ΔT_m was calculated by taking the difference between T_m of the protein with ssDNA and T_m of the control (protein only).

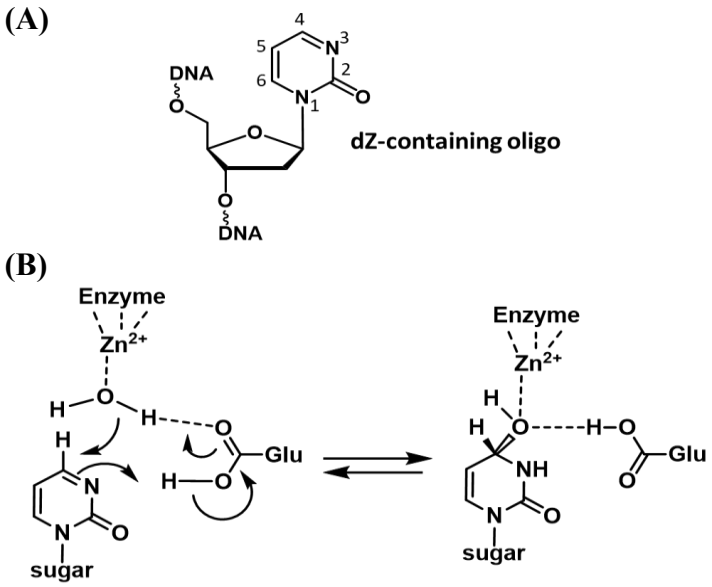


Figure 3.10 Proposed inhibitory mechanism of dZ-containing oligo.
(A) dZ-modified oligonucleotide as potential inhibitor of cytosine deamination, (B) dZ through its conversion into the hydrate and formation of a putative transition-state analogue of cytosine deamination.

These results illustrate that the catalytic glutamate is essential for binding dZ as well as dC-containing oligos. Therefore, the protonation of N3 of the dZ by the catalytic glutamate must be a very significant initial step in the mechanism. The protonation renders C4 to become electrophilic and more prone to the nucleophilic addition of a water molecule (as hydroxide). Therefore, this nucleophilic attack on C4 would transform dZ into a stable tetrahedral transition state of cytosine deamination permitting dZ to remain bound in the A3 active site as we illustrate in Figure 3.10. The potential of dZ-containing oligos to bind to the enzyme with a much higher affinity than the dC-containing oligo suggests that the dZ-containing oligos can inhibit the activity of $A3B_{CTD}$ (later described in chapter 4) and be further used for structural studies (described in chapter 6).

3.4.3 Finding oligonucleotide sequence for further $A3B_{CTD}$-QM-ΔL3-binding studies

Initially, short oligonucleotide sequences containing A3B's preferred deamination motif TCA were designed based on previous studies [36-37, 47, 78, 94, 157, 172]. Structural studies noted that residues surrounding the active site form interactions with nucleotides flanking the preferred dinucleotide deamination motif (5'-TC) [239-240], thereby influencing the A3's substrate recognition and deamination selectivity [175] (described in section 1.6.4). Therefore, several sequences were designed with various nucleotides flanking the TCA motif (see Table 2.13) and assessed using the fluorescence-based thermal shift assay for their capability to bind to $A3B_{CTD}$-QM-ΔL3.

As observed earlier (Figure 3.9), a noticeable increase in the protein's T_m occurs upon binding of an oligonucleotide to the $A3B_{CTD}$-QM-ΔL3 (0.7-2.0 °C, Figure 3.12). Nevertheless, this assay displayed that the substrate-bound enzyme (dC) had a slightly higher change in the melting temperature in comparison to product-bound enzyme (dU), consistent with previous reports of better substrate binding to A3 enzymes [23, 93, 239, 245]. Structural studies have implied that only upon the presence of oligonucleotides with dC that sequence-specific interactions with the protein occur, allowing for the oligonucleotide to fit deeper into the enzyme's active-site pocket, which may justify the stronger preference of dC-containing oligonucleotides [93, 239].

Although wild-type $A3B_{CTD}$ was reported to have a slight (1.4 fold) preference for an adenine in the -2 position (5'-AT<u>C</u>A) over a thymidine (5'-TT<u>C</u>A) [37, 47, 157], $A3B_{CTD}$-QM-ΔL3 prefers the presence of a thymidine in the -2 position (5'-TT<u>C</u>A) in the

oligonucleotide. The A9C-oligo (5'-AATTCAAAA) resulted in the largest ΔT_m of 1.96 ± 0.14 °C, followed closely by the 9merC-oligo (5'-ATTTCATTT) with a ΔT_m of 1.73 ± 0.06 °C, whereas the ATC oligonucleotide containing an adenine in the -2 position (5'-ATCA) resulted in a reduction in the protein's ΔT_m of 1.03± 0.08 °C (Figure 3.12). However, our outcome is consistent with the ssDNA bound $A3B_{CTD}$ ($A3B_{CTD}$-QM-ΔL3-AL1swap-E255A) crystal structure, which showed that an arginine (R211) from loop 1 interacts with the backbone phosphate of a thymidine in the -2 position, whilst retaining the U-shaped DNA binding conformation [239-240] (described in section 1.6.4). Based on this assay, the A9C-oligo had the highest binding affinity for $A3B_{CTD}$-QM-ΔL3. Therefore, this oligonucleotide was selected for further experiments.

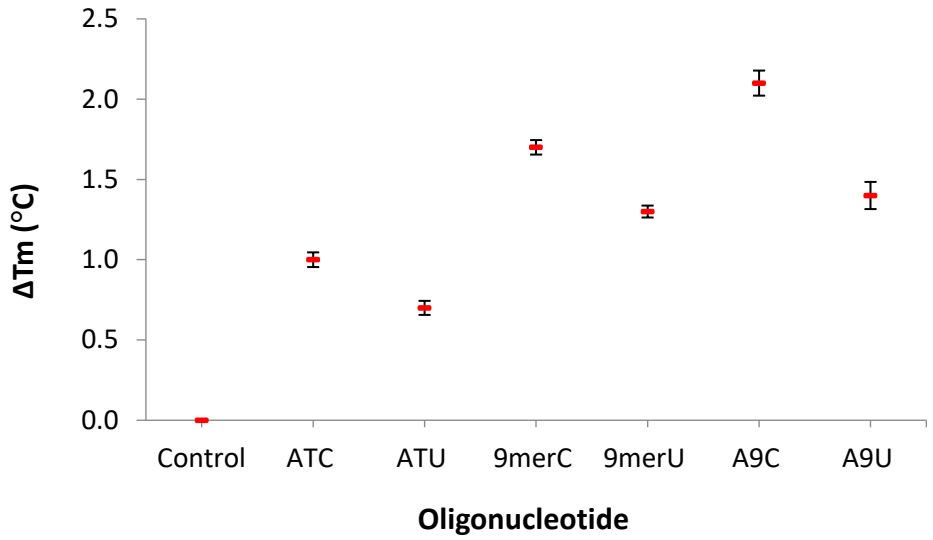


Figure 3.12 Oligonucleotide binding capability to $A3B_{CTD}$ assessed using a thermal shift fluorescense assay.

Change of melting temperature (ΔT_m) of $A3B_{CTD}$-QM-ΔL3 calculated from the difference between the melting temperature (T_m) of the protein in presence of ssDNA and T_m of the protein in the buffer (control). Assays were setup using 20 µM $A3B_{CTD}$-QM-ΔL3 protein along with 100 µM ssDNA oligonucleotides in 50 mM citrate-phosphate pH 5.5 buffer (50 mM citrate-phosphate buffer, 200 mM NaCl, 2 mM β-ME, 200 µM DSS, pH 5.5) containing 10X SYPRO® orange dye (BioRad). Confidence intervals (95%) are shown as error bars.

Moreover, to directly characterise the molecular interaction between $A3B_{CTD}$-QM-ΔL3 and the substrate oligonucleotide isothermal titration calorimetry (ITC) was utilised. ITC hinges on assessment of fluctuations in heat that is released or absorbed upon the occurrence of a binding interaction in a solution. The experiment was setup using 100 µM $A3B_{CTD}$-QM-ΔL3 titrated with 300 µM substrate oligo (A9C-oligo) in the 50 mM

citrate-phosphate pH 5.5 buffer (refer to section 2.4.2). However, under these conditions the determination of a very weak binding interaction between the enzyme and substrate oligonucleotide was difficult. Characterising this weak binding event by ITC would require much higher concentrations of reactants, which is not feasible for the protein as it is inclined to precipitation at increased concentrations.

3.5 NMR structural studies of A3B$_{CTD}$-QM-ΔL3

To gain insight into the molecular mechanisms underlying the binding of ssDNA to the A3B$_{CTD}$-QM-ΔL3 nuclear magnetic resonance (NMR) spectroscopy was used. NMR is a distinctive technique that reveals the atomic structure of proteins in solution and their dynamic behaviours. This technique exploits the fact that certain atomic nuclei, such as ^{1}H, ^{15}N, and ^{13}C isotopes, are magnetic when placed in a magnetic field. One-dimensional proton-NMR (1D ^{1}H-NMR) examines with high sensitivity the naturally abundant proton (^{1}H) isotope nuclei within molecules. However, in comparison to small-molecule compounds, complicated overlapping signals arise in 1D ^{1}H-NMR of proteins due to a high number of protons leading to complex spectra analysis. Therefore, multidimensional (2D and 3D) NMR experiments are used to resolve the overlapping signals. Multidimensional experiments rely on the transfer of magnetisation energy between nuclei, providing specific correlations between nuclei. To enable these experiments, proteins are labelled with isotope nuclei (^{13}C, and ^{15}N) (refer to 2.2.6) to observe chemical shifts in NMR spectra.

3.5.1 Uniform isotopically-labelled A3B$_{CTD}$-QM-ΔL3 protein

Proteins are naturally abundant with proton (^{1}H) isotope nuclei, but to enable multidimensional NMR experiments proteins are isotopically labelled. Proteins are expressed in minimal medium artificially enriched with an isotope nucleus of interest. The A3B$_{CTD}$-QM-ΔL3 variant was prepared in minimal medium enriched with either ^{15}N isotope to produce single-labelled protein (^{15}N-A3B$_{CTD}$-QM-ΔL3) for use in 2D NMR experiments, or with both ^{15}N and ^{13}C isotopes to obtain double-labelled protein (^{15}N,^{13}C-A3B$_{CTD}$-QM-ΔL3) for use in 3D-NMR experiments (refer to 2.2.6).

Initial expression of the isotope-labelled A3B$_{CTD}$-QM-ΔL3 protein in minimal medium led to very poor protein yields, in contrast to expression in the nutrient rich LB medium, which results in high protein yields (Section 3.2 and Figure 3.6). To obtain high protein yields using isotopically labelled minimal media, several expression optimisation

conditions were assessed. We mainly found that induction of protein expression during the early growth phase of cells (OD_{600} between 0.3-0.4) led to a modest improvement in the production of the labelled $A3B_{CTD}$-QM-ΔL3 protein. Additionally, it was noted that a decrease in the pH occurs during cell growth in minimal medium as a consequence of the accumulation of acidic by-products and deficiency of buffering capacity. This pH drop may potentially cause cell stress and consequently the observed decrease in the protein yield as found in previous studies [5, 248]. Therefore, to maintain the buffering capacity of the minimal medium during cell growth, the pH of the minimal medium was adjusted to pH 7.5 with 1 M NaOH before culture inoculation and just before induction of the protein. These adjustments led to a significant increase in the expression of the isotopically labelled $A3B_{CTD}$-QM-ΔL3 variant as shown in lanes 4 and 8-10 in Figure 3.13. Using these conditions, the labelled $A3B_{CTD}$-QM-ΔL3 protein was expressed in a large-scale setup (outlined in section 2.2.6), for use in multidimensional NMR experiments.

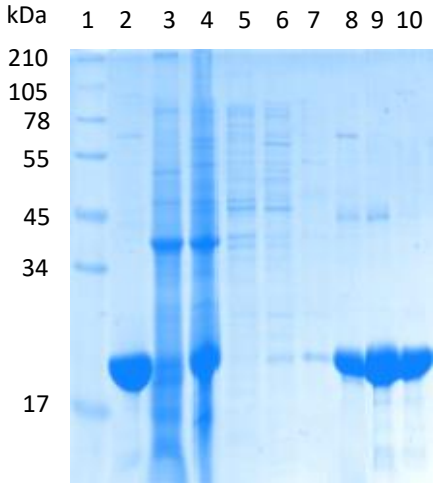


Figure 3.13 Tricine SDS-PAGE (16%) gel of purification of ^{15}N, ^{13}C-$A3B_{CTD}$-QM-ΔL3.

Lane 1: SeeBlue Plus2 Pre-stained protein standard ladder, lane 2: unlabelled $A3B_{CTD}$-QM-ΔL3 as control, lane 3: pellet, Lane 4; supernatant, lanes 5-7: buffer wash fractions *, lanes 8-9: eluted fractions $^{+}$, lane 10: eluted SEC fraction ^{X}. *Wash buffer: 50 mM Tris-HCl pH 7.4, 0.5 M NaCl, 2 mM β-ME, with either 10 mM, 20 mM, or 50 mM imidazole. $^{+}$Elution buffer: 50 mM Tris-HCl pH 7.4, 0.5 M NaCl, 2 mM β-ME, with either 200 mM, or 300 mM imidazole. ^{X}SEC buffer: 50 mM Tris-HCl pH 7.4, 0.5 M NaCl, 2 mM β-ME (see section 2.2.6).

3.5.2 2D ^{1}H,^{15}N-HSQC spectrum of $A3B_{CTD}$-QM-ΔL3

At first, a 1D ^{1}H-NMR spectrum of $A3B_{CTD}$-QM-ΔL3 was recorded to test the proton peak-dispersion pattern and establish the optimum protein concentration required for

high-resolution NMR experiments. Next, a 2D $^{1}H,^{15}N$-heteronuclear single quantum coherence ($^{1}H,^{15}N$-HSQC) spectrum was acquired to assess the overall quality of the protein as described in section 2.6.1. Each cross peak (resonance) represents the correlation between the proton nuclei directly bound to the nitrogen nuclei (amide bond) and can be assigned to a particular amino acid residue in the protein.

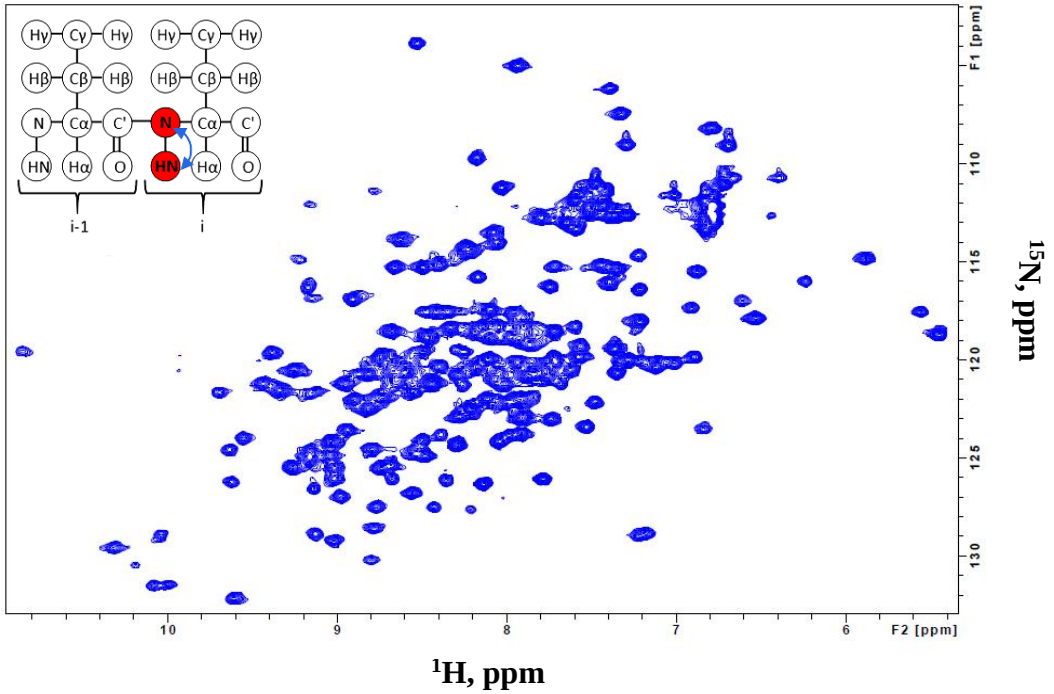


Figure 3.14 $^{1}H,^{15}N$-HSQC of the ^{15}N-A3B$_{CTD}$-QM-ΔL3.

700 MHz NMR spectrum of 350 µM of ^{15}N-A3B$_{CTD}$-QM-ΔL3 in high salt pH 7.4 buffer (500 mM NaCl, 50 mM Tris-HCl pH 7.4, 5 mM β-ME, 10% D_2O) at 298 K. Each peak in the spectrum represents a bound N-H pair corresponding to an amino acid residue of the protein. Axes correspond to chemical shifts of ^{1}H and ^{15}N in ppm. Map in top left corner represents pattern of ^{1}H-^{15}N polarisation.

We expected to observe 187 amide proton resonances, which constitute the amino acids of A3B$_{CTD}$-QM-ΔL3 excluding 9 prolines (which do not contain amide protons), along with six amino acids of the histidine-tag (His_6). The spectrum showed that the protein contained the approximate correct number of cross peaks (Figure 3.14). In addition, the spectrum indicated that ^{15}N-A3B$_{CTD}$-QM-ΔL3 was well-folded, with generally distinguished and well-dispersed resonances, apart from several overlapped resonances shown in Figure 3.14. To be able to assign the resonances of the HSQC spectrum 3D-NMR experiments were employed.

3.5.3 Assignment of A3B$_{CTD}$-QM-ΔL3 using 3D-NMR experiments

At the time of this study only the single-domain A3A NMR structure was available [36]. Based on the 89 % sequence identity between A3A and A3B$_{CTD}$ seen in Figure 3.15, the A3A assignment was utilised as a starting point for the assignment of the A3B$_{CTD}$ variant resonances.

```
A3A       MEASPASGPRHLMDPHIFTSNFNNG---IGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYGRHA
A3B_CTD   --------------PDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTWVLMDQHMGFLCNEAKNLLCGFYGRHA
                       *  ** ****    + * +************** * **** *** *+**************

A3A       ELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRD
A3B_CTD   ELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRD
          ********************************************************************************

A3A       AGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN
A3B_CTD   AGAQVSIMTYDEFEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQNQGN
          ************++******  ***********+********************
```

Figure 3.15 Sequence alignment of A3A and A3B$_{CTD}$.

Alignment of the amino acids of the single domain A3A and the C-terminal domain of A3B using BLAST (blastp suite-2sequences). Asterisks indicate an identical match, plus signs indicate conservative substitutions.

Nonetheless, to enable the assignment of the protein backbone, numerous triple-resonance (3D) NMR experiments were conducted (detailed in Table 8.1). These experiments provide sequential connectivities (^{1}H, ^{13}C, ^{15}N) between residues by directly transferring magnetisation over the peptide bond and are named accordingly based on the magnetisation path between the nuclei. Experiments were carried out using purified double-labelled ^{15}N, ^{13}C-A3B$_{CTD}$-QM-ΔL3 as described in the method 2.2.6. Assignment was conducted by Dr. Elena Harjes using the collected 3D-NMR spectra and based on the published A3A assignment.

HNCO and HN(CA)CO NMR spectra were first collected to obtain CO chemical shifts that aid in the prediction of the secondary structure and the protein backbone. The highly sensitive HNCO experiment produced a relatively good spectrum; however, the HN(CA)CO spectrum was unable to be collected. As both these spectra were required in conjunction, the residue connectivities were unable to be resolved. Next, an HNCA spectrum was acquired resulting in a comparatively good spectrum. An HNCA experiment transfers the magnetisation between HN and CA of the same and preceding amino acid and therefore is useful for the assignment of the protein's backbone. To further resolve degenerate resonance frequencies two complementary experiments, HNCACB

and CBCA(CO)NH, were recorded to provide sequential connectivities of the resonances. However, owing to the long acquisition period (over 5 days) some protein precipitation occurred resulting in poor quality spectra. Despite this, these spectra were still able to provide some information that supported the assignment of the protein backbone.

The combination of the triple resonance NMR experiments and NOESY spectrum with the short mixing time facilitated the assignment of nearly all the backbone amide signals (175 out of the 187 observable), as well as most of the hydrogens linked to the alpha carbons and side chain (>70%) of A3B$_{CTD}$-QM-ΔL3 as shown in Figure 3.16 and listed in Table 8.1.

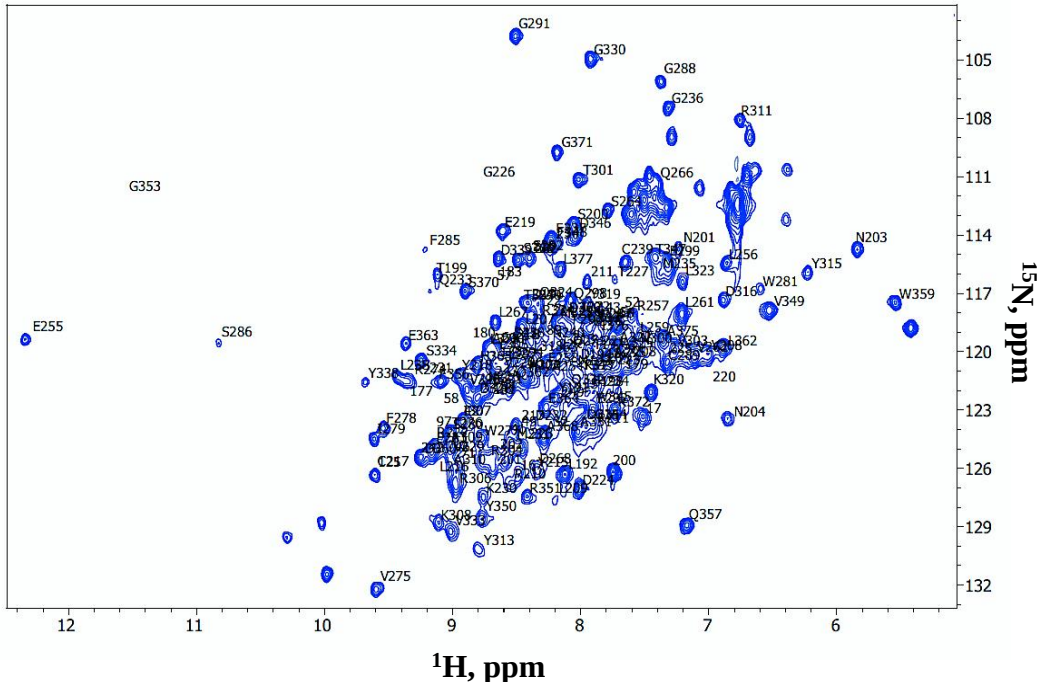


Figure 3.16 A3B$_{CTD}$-QM-ΔL3 assigned ^{1}H,^{15}N-HSQC spectrum.

3D-NMR experiments conducted for assignment of the spectrum using 350 µM of ^{15}N,^{13}C-A3B$_{CTD}$-QM-ΔL3 in high salt pH 7.4 buffer (500 mM NaCl, 50 mM Tris-HCl pH 7.4, 5 mM β-ME, 10% D_2O) at 298 K. Assignments of the ^{1}H,^{15}N-HSQC spectrum resonances indicated with the name and number of the residue.

During this study, the NMR solution state structure of A3B$_{CTD}$-Wt [37] and the crystal structure of an A3B$_{CTD}$ variant [240] were published. Therefore, we did not proceed with structure determination of our ligand-free A3B$_{CTD}$-QM-ΔL3, instead choosing to focus on examining the ssDNA interactions to the A3B$_{CTD}$-QM-ΔL3 protein.

3.5.4 Chemical shift perturbations to identify ssDNA binding site in A3B$_{CTD}$-QM-ΔL3

As chemical shifts are sensitive to the local electronic environment of the nucleus, chemical shift perturbation experiments can be utilised to identify the location and affinity of a ligand for a protein. Exploiting $^{1}H,^{15}N$-HSQC experiments, chemical shift changes can be monitored upon the addition of increasing amounts of ligand to protein. These changes give an indication of residues that may be involved in interactions between the ligand and protein, allowing for the identification of the binding site.

In attempt to study the binding of ssDNA to A3B$_{CTD}$-QM-ΔL3, samples were setup in the previously mentioned 50 mM citrate-phosphate pH 5.5 buffer (50 mM citrate-phosphate buffer, 200 mM NaCl, 2 mM β-ME, and 200 µM DSS, pH 5.5, 10% D_2O) to entice interactions between the protein and ssDNA. Initially, a $^{1}H,^{15}N$ HSQC spectrum of 250 µM ^{15}N-A3B$_{CTD}$-QM-ΔL3 was compared to spectra of the protein titrated with varying amounts of the A9C-oligo (5'-AATTCAAAA) in the pH 5.5 buffer, as described in section 2.4.3. However, assessment of the spectra found that upon titration of the protein with increasing amounts of oligonucleotide the majority of the peaks shifted in one direction. This was presumed to be caused by the addition of large amounts of DNA, leading to changes in the pH of the solution as a consequence of the negative charges on the DNA phosphates. Hence, the protein and oligonucleotide were dialysed in a matched buffer, and the pH of the sample was monitored following data acquisition to ensure that changes in cross-peaks in the NMR spectra occur only as a result of the addition of ssDNA to the protein.

Using this method, we trialled titration of the A3B$_{CTD}$-QM-ΔL3 with A9C-oligo at a 1 to 5 and a 1 to 20 ratio. This resulted in a few signals shifting (Figure 3.17), which was a good indication that the ssDNA was bound to the protein in a pH-controlled environment. Determining the surface where the ssDNA binds to the protein and its binding affinity required further titrations with a range of substrate concentrations. However, during this study the structure of ssDNA bound to an inactive A3B$_{CTD}$ variant (A3B$_{CTD}$-QM-ΔL3-AL1swap-E255A) was published [37, 239]. So, the study then focused on the evaluation of A3B inhibitors. Nevertheless, the assignment obtained for our A3B$_{CTD}$-QM-ΔL3 variant can be used to study interactions of other small molecules and binding partners of A3.

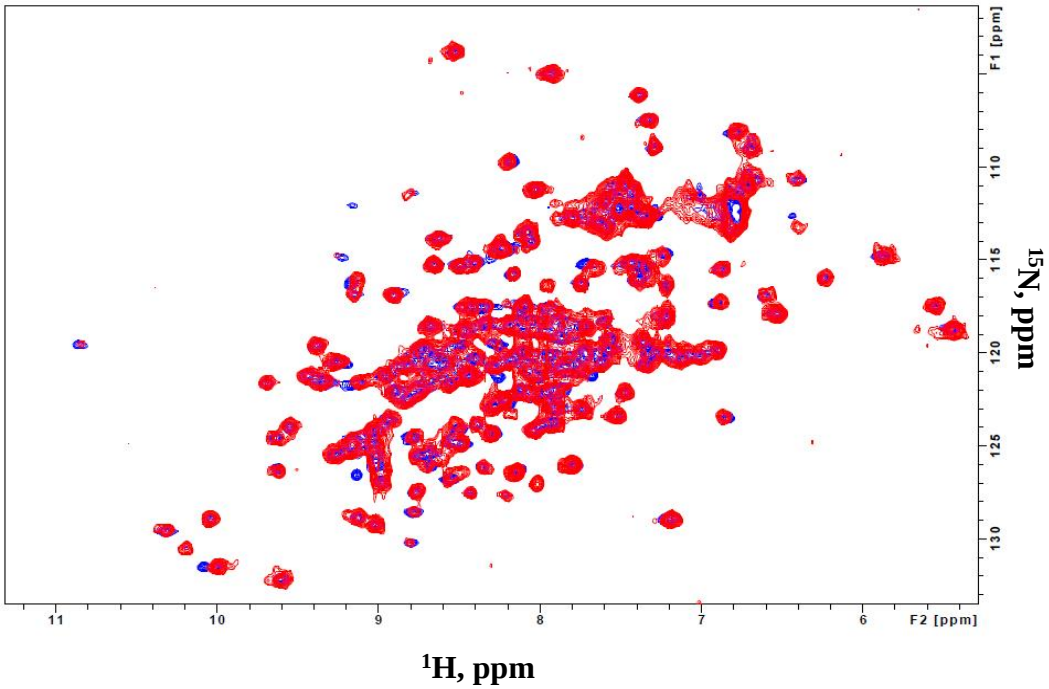


Figure 3.17 ^{1}H,^{15}N-HSQC spectra overlay of a ^{15}N-A3B$_{CTD}$-QM-ΔL3 spectrum against a ^{15}N-A3B$_{CTD}$-QM-ΔL3 titrated with A9C-oligo.

^{1}H,^{15}N-HSQC 700 MHz spectrum of 250 µM ^{15}N-A3B$_{CTD}$-QM-ΔL3 (blue) titrated with A9C-oligo (5'-AATTCAAAA) at a 1:20 ratio (red) (protein to oligonucleotide) in a 50 mM citrate-phosphate pH 5.5 buffer (50 mM citrate-phosphate buffer, 200 mM NaCl, 2 mM β-ME, and 200 µM DSS, pH 5.5, 10% D_2O) at 298 K.

3.6 Conclusions

We first purified catalytically inactive (on the time scale of our experiments) A3B$_{CTD}$ variants (A3B$_{CTD}$-QM-ΔL3 and A3B$_{CTD}$-QM-ΔL3-E255A) for the assessment of the binding affinity of ssDNA sequences and to prevent complications of cytosine deamination during a reaction. One key finding indicated that the catalytic glutamate (E255) within the A3 enzymes' active site was critical for binding ssDNA, in particular dZ- or dC-containing oligos. Furthermore, the presence of this glutamate resulted in a higher binding affinity towards the dZ cytosine analogue over the dC substrate, indicating that dZ would mimic the transition state or an intermediate in the cytosine deamination reaction and potentially inhibit the activity of the A3B$_{CTD}$. From these results, we proposed the A3 inhibition mechanism by dZ, whereby the essential E255 would initiate the reaction by protonating N3 of dZ resulting in an electrophilic C4 which can be attacked by a water molecule. This effectively would convert the dZ into a rigid tetrahedral transition state of cytosine deamination that would bind tighter in the the active

site than the substrate, as observed in our thermal shift experiments, and effectively inhibit the catalytic activity.

Therefore, the $A3B_{CTD}$-QM-ΔL3 variant with the intact catalytic glutamate was used for out binding studies. Using NMR spectroscopy, we assigned the resonances of the ligand-free $A3B_{CTD}$-QM-ΔL3, but as ligand-free $A3B_{CTD}$ structures [37, 240] became available we focused on identifying the binding site and affinity of substrate oligonucleotides to our inactive $A3B_{CTD}$ variant using chemical shift perturbations. From this experiment we identified that the ssDNA bound to the protein, though further assessment using substrate ssDNA was not continued as the structures of inactive $A3B_{CTD}$ in complex with ssDNA were published [37, 239]. Using the thermal shift assay and ITC, we found that the inactive $A3B_{CTD}$ had a very weak binding affinity for ssDNA. As binding was not the best indicator of the inhibitory potential of the cytosine analogue-containing ssDNA using an inactive $A3B_{CTD,}$ the study moved its focus to directly evaluate the inhibition of catalytically active $A3B_{CTD}$ by our modified-oligonucleotides using an ^{1}H-NMR based assay [94].

4. Inhibition of APOBEC3B deamination activity using modified ssDNA

4.1 Introduction

The majority of the work reported in this chapter was incorporated in the following two published journal articles:

1) Kvach, M. V., Barzak, F. M., Harjes, S., Schares, H. A., Jameson, G. B., Ayoub, A. M., Moorthy, R., Aihara, H., Harris, R. S., Filichev, V., Harki, D. A., and Harjes, E. Inhibiting APOBEC3 activity with single-stranded DNA containing 2'-deoxyzebularine analogues. *Biochemistry* 2019, 58 (5), 391- 400.
doi: 10.1021/acs.biochem.8b00858. [135].

   Joint first author. In this article, I expressed and purified all the $A3B_{CTD}$ variant enzymes ($A3B_{CTD}$-QM-ΔL3, $A3B_{CTD}$-QM-ΔL3-E255A, $A3B_{CTD}$-QM-ΔL3-AL1swap, and $A3B_{CTD}$-DM). Applying these enzymes, I performed thermal shift assays and ITC experiments to assess the binding affinity of dC- and dZ-containing-ssDNA towards the $A3B_{CTD}$ enzymes. I conducted ^{1}H-NMR based activity assays, characterising the kinetic deamination activity of the $A3B_{CTD}$ variants on substrate oligonucleotide (Oligo-2). In addition, I characterised the $A3B_{CTD}$ enzymes' inhibition mode and inhibition potency against our dZ-containing oligonucleotide using a ^{1}H-NMR based inhibition assay. I performed all analysis of the derived data.

2) Kvach, M. V., Barzak, F. M., Harjes, S., Schares, H. A., Kurup, H. M., Jones, K. F., Sutton, L., Donahue, J., D'Aquila, R. T., Jameson, G. B., Harki, D. A., Krause, K. L., Harjes, E., and Filichev V. Differential inhibition of APOBEC3 DNA-mutator isozymes by fluoro- and non-fluoro-substituted 2'-deoxyzebularine embedded in single-stranded DNA. *ChemBioChem* 2020, 21 (7), 1028-1035. doi.org/10.1002/cbic.201900505 [136].

   Joint first author. In this article, I expressed and purified the catalytically active $A3B_{CTD}$ variant enzymes ($A3B_{CTD}$-AL1 and $A3B_{CTD}$-DM). I quantitatively screened our modified oligonucleotides (3dadZ-oligo, 3dadU-oligo, dZ-oligo, and 5FdZ-oligo) against the deamination activity of $A3B_{CTD}$-AL1 to assess their inhibitory potential by using the ^{1}H-NMR based inhibition assay. From this screen the 5FdZ-oligo was identified as a new inhibitor, for which I characterised its inhibition constant against $A3B_{CTD}$-AL1 using the ^{1}H-NMR based inhibition assay. I preformed all analysis of the derived data.

The work reported in this chapter was also presented in the conference listed below.

Barzak, F. M., Kvach, M. V, Harjes S., Jameson G.B, Aihara, H., Harris R. S., Harki, D. A., Filichev, V. V., Harjes, E. DNA-based inhibitors of the human APOBEC3B DNA cytosine deaminase. Invited oral presentation at ComBio2018 held between the 23rd - 26th of September 2018 at the International Convention Centre, Darling Harbour, Sydney, Australia.

Earlier studies reported several small molecule inhibitors of A3G [151, 196]. However, so far no selective inhibitors have been reported for the tumour-promoting A3B (and A3A). Therefore, modified oligonucleotides, in which cytosine analogues known to inhibit the closely related CDA were incorporated in place of the target dC, were synthesised and evaluated for their inhibitory potential of the $A3B_{CTD}$.

This chapter first describes the production and purification of catalytically active $A3B_{CTD}$ constructs. The Michaelis-Menten constants of the rate of deamination on a substrate oligonucleotide by active $A3B_{CTD}$ variants were then characterised using a 1D-NMR activity assay [79, 94]. Using this assay, chemically modified ssDNAs were screened for their inhibitory potential on $A3B_{CTD.}$ Two modified oligos showed significant inhibition of the $A3B_{CTD}$ catalytic activity. Qualitative revaluation of these ssDNA-containing transition-state analogues against active $A3B_{CTD}$ variants showed that these modified oligos had low micromolar inhibition constants. These oligonucleotides constitute the first substrate-like A3 inhibitors.

4.2 Protein expression and purification of $A3B_{CTD}$ variants: $A3B_{CTD}$-QM-ΔL3-AL1swap and $A3B_{CTD}$-DM

To investigate the potency of the chemically modified ssDNA-based inhibitors against the A3B enzyme, catalytically active $A3B_{CTD}$ enzymes were required. Since wild-type $A3B_{CTD}$ was unable to be expressed and purified effectively, catalytically active $A3B_{CTD}$ derivatives were utilised, which were described by Shi et al [240].

The $A3B_{CTD}$-QM-ΔL3-AL1swap variant (for simplicity referred to as $A3B_{CTD}$-AL1 in this chapter) contained four substitution mutations (F200S, W228S, L230K, F308K), the removal of loop 3, and the transplant of loop 1 from A3A, as shown in Figure 4.1. The SUMO-$A3B_{CTD}$-DM variant had two mutations (L230K and F308K) away from the active-site displayed in Figure 4.1 and had a His_6-SUMO fusion protein (not shown in Figure 4.1) to aid the expression and solubility of the protein.

```
A3BCTD                      ------EILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTWVLMDQHMGFLCN
A3BCTD-DM                   ------EILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTWVKMDQHMGFLCN
A3BCTD-QM-ΔL3-AL1swap       ------EILRYLMDPDTFTSNFNNG---IGRHKTYLCYEVERLDNGTSVKMDQHMGFLCN
                                  *    *********             * *

A3BCTD                      EAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQEN
A3BCTD-DM                   EAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQEN
A3BCTD-QM-ΔL3-AL1swap       E--------SGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQEN
                             *********

A3BCTD                      THVRLRIFAARIYDY-DPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPW
A3BCTD-DM                   THVRLRIKAARIYDY-DPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPW
A3BCTD-QM-ΔL3-AL1swap       THVRLRIKAARIYDY-DPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPW
                                    *

A3BCTD                      DGLEEHSQALSGRLRAILQ----
A3BCTD-DM                   DGLEEHSQALSGRLRAILQ----
A3BCTD-QM-ΔL3-AL1swap       DGLEEHSQALSGRLRAILQ----
```

Figure 4.1 A3B$_{CTD}$ variants amino acid sequence used for deamination assays.

Amino acid sequence of the wild-type A3B$_{CTD}$ residues 187 to 388. A3B$_{CTD}$ variants are A3B$_{CTD}$-DM (L230K and F308K) and A3B$_{CTD}$-QM-ΔL3-AL1swap (F200S, W228S, L230K, and F308K, loop 3 deletion, loop 1 from A3A) [240]. Asterisks represents the variations and modifications in comparison to the wild-type A3B$_{CTD}$.

4.2.1 Expression of A3B$_{CTD}$ variants (A3B$_{CTD}$-QM-ΔL3-AL1swap and A3B$_{CTD}$-DM)

The active A3B$_{CTD}$ plasmid variants, pET24a-A3B$_{CTD}$-QM-ΔL3-AL1swap and pE-SUMO-A3B$_{CTD}$-DM (acquired from Assoc. Prof. Hideki Aihara and Prof. Reuben Harris, University of Minnesota) [240], were expressed in a similar manner to our inactive A3B$_{CTD}$ proteins described earlier (see section 3.2). Briefly, the plasmids were transformed into chemically competent *E. coli* cells, before being isolated, and sequenced confirming inserts (refer to 2.2). The isolated plasmids were then visualised on an agarose gel (Figure 4.2).

The presence of a band in lane 2 that migrated at around 6.4 kb for the putative SUMO-A3B$_{CTD}$-DM plasmid, and in lane 3 the band present at around 6 kb for the pET24a-A3B$_{CTD}$-QM-ΔL3-AL1swap plasmid were consistent with the overall size of the plasmid with the A3B$_{CTD}$ insert (pET24a plasmid ~ 5.4 kb, pE-SUMO plasmid ~ 5.7 kb, A3B$_{CTD}$ insert ~ 0.6 kb).

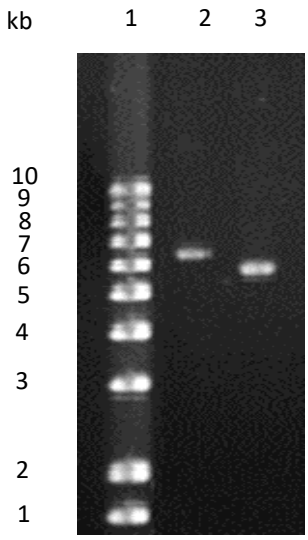


Figure 4.2 Agarose gel (1%) of the isolated A3B$_{CTD}$ plasmid variants.

Lane 1: 1kb plus DNA standard ladder, lane 2: pE-SUMO-A3B$_{CTD}$-DM plasmid isolate, lane 3: pET24a- A3B$_{CTD}$-QM-ΔL3-AL1swap plasmid isolate. Gel run at 100 V for approximately 50 min.

These A3B$_{CTD}$ plasmids were first expressed in a BL21(DE3) *E. coli* strain in a small-scale trial as described in the previous chapter (see section 3.2.2). However, during the incubation period the cells ceased to grow exponentially (mid-log phase), which was attributed to be a consequence of cell toxicity caused by leaky expression of the catalytically active A3B$_{CTD}$ leading to mutagenesis of the bacterial DNA. Therefore, the BL21(DE3) *E. coli* derivatives (OverExpress; C41(DE3), C43(DE3), C41(DE3) pLysS, and C43(DE3) pLysS cells) that tolerate toxic proteins by minimising background expression were screened. We found that the C41(DE3) pLysS *E. coli* strain was the most suitable at expressing the active A3B$_{CTD}$.

Further optimisations were conducted to improve the protein solubility. First, the incubation temperature was lowered from 37 °C to 25 °C, then protein expression was induced at an earlier OD_{600} of 0.3 to limit solubility issues. It can be seen on the SDS-PAGE gel that only upon IPTG induction does an additional band appear implying that the A3B$_{CTD}$ constructs are expressed. A band present with a size of approximately 22 kDa corresponded to the A3B$_{CTD}$-AL1 protein, seen in lane 3 in Figure 4.3A. The dense band corresponding to the fusion SUMO-A3B$_{CTD}$-DM protein migrated at around 38 kDa (6XHis-SUMO is 16 kDa and A3B$_{CTD}$-DM is 22 kDa) (observed in Figure 4.3B).

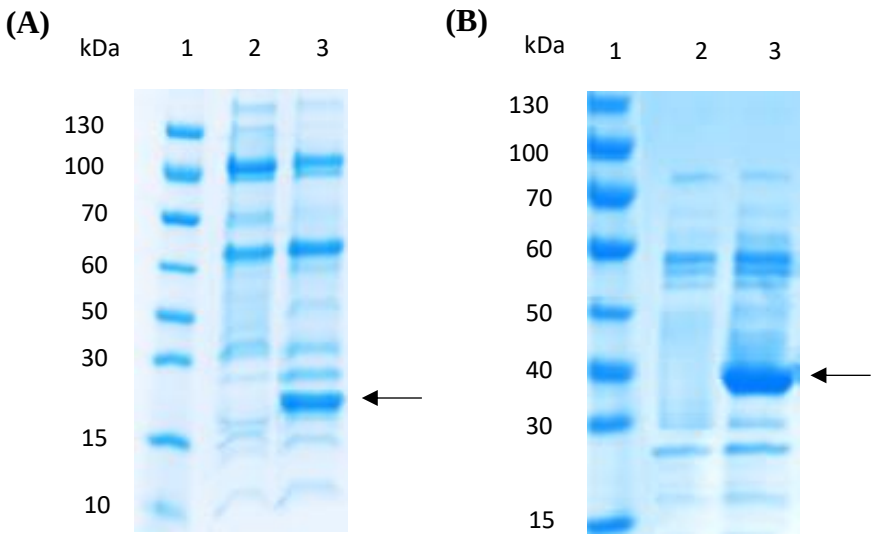


Figure 4.3 Tricine SDS-PAGE gels (16%) of the protein expression of the active A3B$_{CTD}$ constructs.

(A) A3B$_{CTD}$-QM-ΔL3-AL1swap, (B) SUMO-A3B$_{CTD}$-DM Lane 1: Blue Prestained protein standard ladder, lane 2: uninduced cell culture, lane 3: induced cell culture with 0.5 mM IPTG. Arrows point to A3B$_{CTD.}$ Gel run at 200 V for approximately 50 min.

In addition to the appearance of bands following induction, the proteins' identities were validated using mass spectrometry as described in an earlier chapter (see section 3.2.2). Noticeably, the SUMO-A3B$_{CTD}$-DM protein was expressed in higher concentrations in comparison to the A3B$_{CTD}$-AL1 protein as seen by a higher density of the protein band (Figure 4.3), which may be associated with the slightly lower deamination activity of the A3B$_{CTD}$-DM protein in comparison to the other construct, as reported [240].

4.2.2 Purification of A3B$_{CTD}$ variants (A3B$_{CTD}$-QM-ΔL3-AL1swap and A3B$_{CTD}$-DM)

Using the optimised conditions discussed earlier (refer to section 3.2.2), large-scale productions of the histidine-tagged active A3B$_{CTD}$ proteins (SUMO-A3B$_{CTD}$-DM and A3B$_{CTD}$-QM-ΔL3-AL1swap) were conducted. The collected protein fractions from IMAC were visualised on SDS-PAGE gels, showing that the proteins of interest (A3B$_{CTD}$) had eluted off the column, as indicated by the presence of the correct-sized band (His$_6$-A3B$_{CTD}$-QM-ΔL3-AL1 ~ 21 kDa, His$_6$-SUMO-A3B$_{CTD}$-DM ~ 38 kDa) (Figure 4.4). However, background contaminant proteins eluted off the column along with the A3B$_{CTD}$ proteins following IMAC; therefore, further purification was employed using size exclusion chromatography.

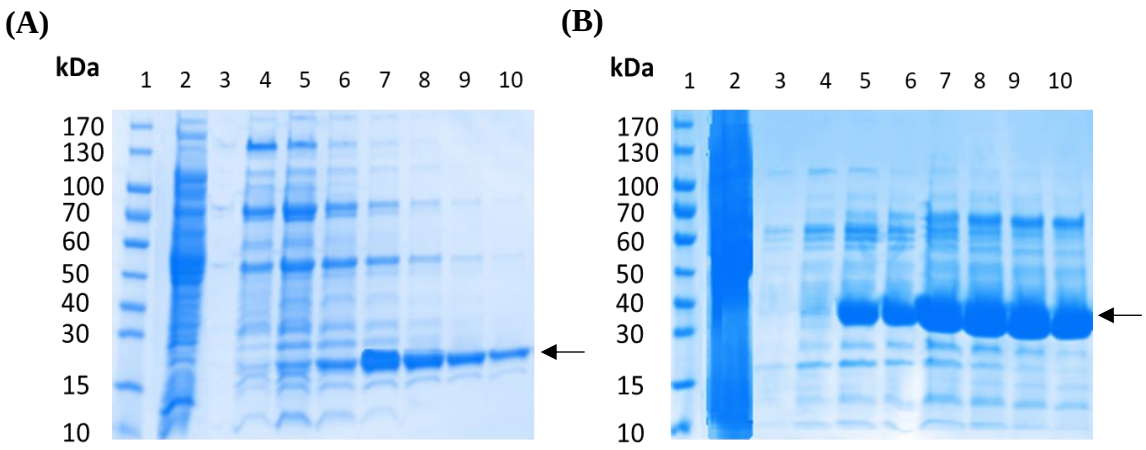


Figure 4.4 Tricine SDS-PAGE gels (16%) of IMAC protein purification of the A3B$_{CTD}$ variants. (A) A3B$_{CTD}$-QM-ΔL3-AL1swap, (B) SUMO-A3B$_{CTD}$-DM. Lane 1: Blue Prestained protein standard ladder, lane 2: supernatant flow-through, lanes 3-4: buffer wash fractions *, lanes 5-10: eluted fractions $^{+}$. *Wash buffer; 50 mM Tris-HCl pH 7.4, 0.5 M NaCl, 2 mM β-ME, with 10 or 50 mM imidazole. $^{+}$Eluting buffer: 50 mM Tris-HCl pH 7.4, 0.5 M NaCl, 2 mM β-ME, with 200 to 500 mM imidazole. Gel run at 200 V for approximately 50 min.

The A3B$_{CTD}$ fractions from IMAC were combined and run through a Superdex 75 10/300 GL column to purify the A3B$_{CTD}$ protein by means of size. The SUMO-A3B$_{CTD}$-DM SEC profile showed a major peak with a maximum around 77 mL eluting off the column with several smaller peaks eluting off earlier. Assessment of the fractions collected from SEC-FPLC of the major peak (peak maximum 77 mL) on an SDS-PAGE gel showed a very dense 38 kDa band had eluted corresponding to the size of the SUMO-A3B$_{CTD}$-DM protein, with minor background contaminants (see Figure 4.4B). After SEC, the fractions containing the most pure SUMO-A3B$_{CTD}$-DM fractions were pooled together and incubated with SUMO protease (ULP1) to cleave off the His$_6$-SUMO tag as described in section 2.2.5.1. Following cleavage, the sample was purified using IMAC to remove the His-SUMO, which binds to the Ni^{2+}-NTA column. The A3B$_{CTD}$-DM protein then eluted directly off the column, which was confirmed by the size of a 22 kDa protein as visualised by Tricine SDS-PAGE (Figure 8.1).

We noticed that upon purification of the A3B$_{CTD}$-AL1 variant two major peaks eluted off the SEC column. The first peak eluting off the column (after 74 mL, Figure 4.5A) was identified to be a dimeric form of the A3B$_{CTD}$-AL1 protein (further discussed in section 6.3), while the peak corresponding to the monomeric A3B$_{CTD}$-AL1 form eluted after 90 mL of buffer (see Figure 4.5A). The monomeric A3B$_{CTD}$-AL1 fractions were collated then run through the SEC column once again ensuring that a homogeneous monomeric

form was isolated. The purified A3B$_{CTD}$ variants were then used in NMR-based activity and inhibition assays.

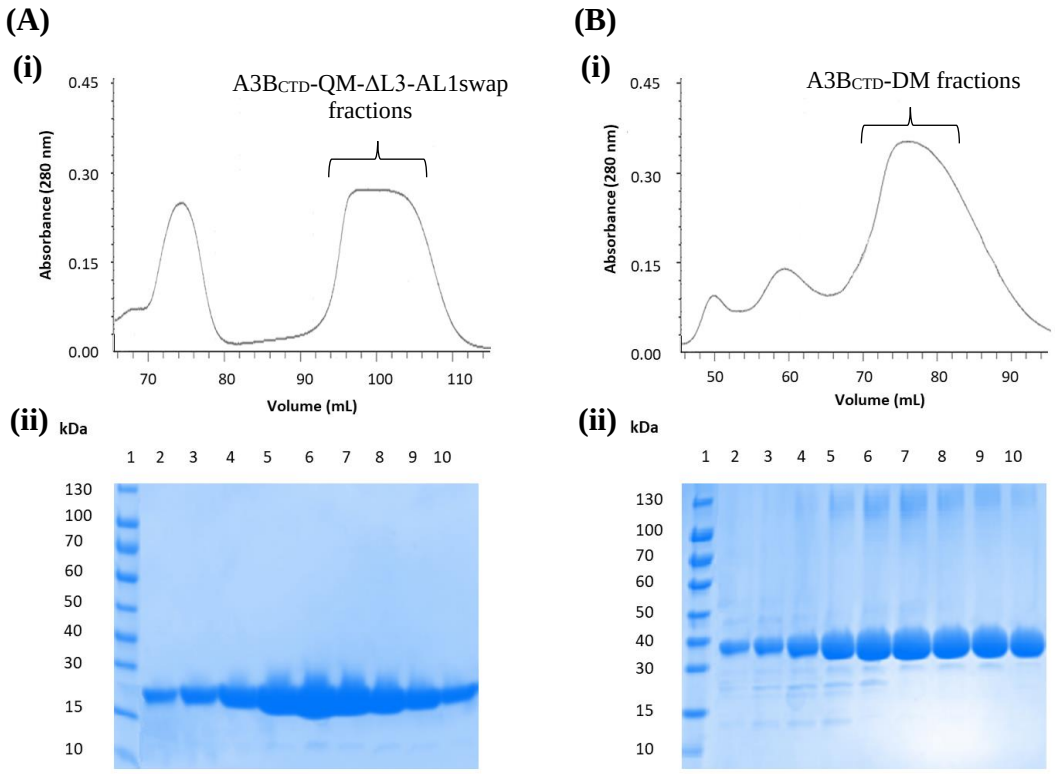


Figure 4.5 SEC protein purification of the active A3B$_{CTD}$ variants.
(A) A3B$_{CTD}$-QM-ΔL3-AL1swap (i) SEC profile, (ii) SDS-PAGE (16%) gel with SEC fractions between 90 – 110 mL, (B) SUMO-A3B$_{CTD}$-DM, (i) SEC profile, (ii) SDS-PAGE (16%) gel with SEC fractions between 70 – 90 mL. SEC buffer is 50 mM Tris-HCl, pH 7.4, 0.5 M NaCl, 2 mM β-ME.

4.3 Evaluation of the catalytic activity of A3B$_{CTD}$-AL1 and A3B$_{CTD}$-DM using an ^{1}H-NMR-based activity assay

4.3.1 A3B$_{CTD}$-AL1 and A3B$_{CTD}$-DM variants catalytic activity

The A3B$_{CTD}$-AL1 and A3B$_{CTD}$-DM variants were previously stated to be catalytically active using a fluorescence-based deaminase activity assay [240]. However, to quantitatively assess the catalytic activity of the purified A3B$_{CTD}$ variants (A3B$_{CTD}$-AL1 and A3B$_{CTD}$-DM) on a dC-containing oligonucleotide, an established 1D-^{1}H-NMR-based activity assay (previously described in section 3.3) was used [79, 94]. However, in place of the A9C-oligo, the 9merC-oligo (from now on referred to as the TCA-oligo) was used for our kinetic studies (see Table 2.13), as it allowed better dispersion of the dC and dU resonances in the ^{1}H-NMR spectra (Figure 4.6), and additionally was shown using thermal shift assay to be capable of binding to our A3B$_{CTD}$-QM-ΔL3 variant (see Figure

3.12). Initial experiments performed under the same conditions as described in section 3.3 with 1 mM TCA-oligo along with 100 µM $A3B_{CTD}$ enzyme under 50 mM citrate-phosphate, pH 7.5 buffer. This assay indicated that these $A3B_{CTD}$ variants were catalytically active, consistent with the aforementioned study [239], although under these conditions the dC of the oligo was rapidly converted to dU (< 5 minutes). Therefore, to examine the deamination of dC to dU qualitatively and quantitatively the concentration of the enzymes were adjusted. We determined that a concentration of 50 nM of $A3B_{CTD}$-AL1 or 2 µM of $A3B_{CTD}$-DM was suitable to observe deamination of the TCA-oligo in a timely manner (several hours), as displayed in Figure 4.6A.

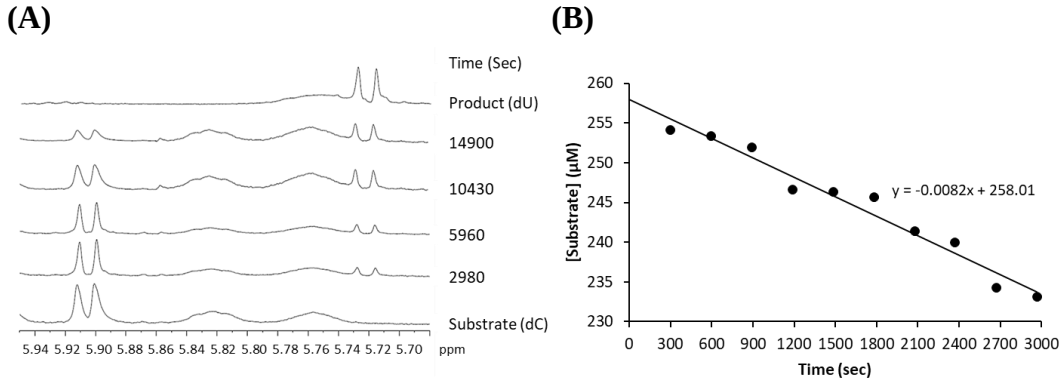


Figure 4.6 Time-resolved deamination of 5'-ATTTCATTT by $A3B_{CTD}$ and assessment of the initial speed of the reaction monitored by ^{1}H-NMR.

(A) A series of ^{1}H spectra was recorded at different time points during the deamination reaction of 350 µM 5'-TCA oligo by 50 nM $A3B_{CTD}$-AL1 in 50 mM citrate-phosphate, pH 5.5 buffer (50 mM citrate-phosphate buffer, 200 mM NaCl, 2 mM β-ME, 200 µM DSS, pH 5.5, with 10% D_2O) monitored over 5 hours. The decrease in the H5 dC proton signal (5.91 ppm) was concurrent with an increase in the H5 dU proton signal (5.73 ppm). (B) The H5 proton signal of dC was integrated, calibrated against a standard citrate buffer signal, and plotted against the initial time of the reaction (~3000 sec) to derive the initial rate of deamination, which was then fitted with a linear regression. ^{1}H-NMR spectra were recorded on a 700 MHz spectrometer every 298 sec at 298 K.

To further optimise the rate of enzyme activity a pH screen was conducted, and the initial deamination rate was calculated. The assay was set up using 500 µM TCA-oligo along with either 50 nM of $A3B_{CTD}$-AL1 or 2 µM $A3B_{CTD}$-DM in the formerly mentioned citrate buffer over a pH range from 4.5 - 7.5. We observed that our A3 variants were catalytically active between pH 4.5–7.5, and therefore would support activity and inhibition of wild-type A3B in physiological pH at around 7 [37]. However, to kinetically characterise these enzymes, use of the same but suboptimal pH is not feasible in NMR assays, as each variant has a distinct optimal pH condition (Figure 4.7). We found that

our A3B$_{CTD}$-DM variant was most active (in the range tested) at pH 7.5 buffer, similar to that reported for wild-type A3B$_{CTD}$ [37]. Meanwhile, the A3B$_{CTD}$-AL1 variant was approximately 6-fold more catalytically active in the more acidic pH 5.5 buffer as shown in Figure 4.7 and Table 4.1. This difference is the result of the transfer of the A3A loop 1 into the A3B$_{CTD}$ essentially gaining characteristic properties of A3A as reported in the literature [37, 107]. Therefore, these pH conditions resulted in an optimum deamination rate for each A3B$_{CTD}$ variant and were selected for our NMR-based activity and inhibition assays.

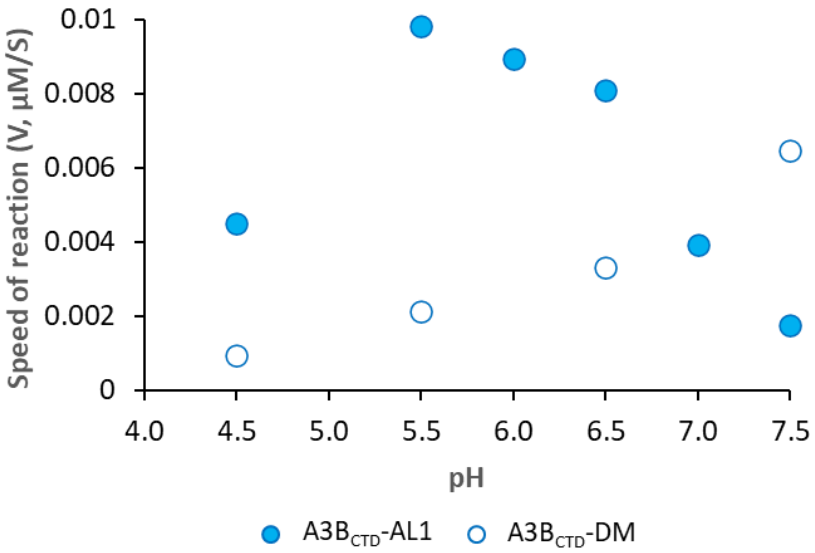


Figure 4.7 pH dependence on the rate of deamination of 5'-ATTTCATTT for A3B$_{CTD}$-AL1 and A3B$_{CTD}$-DM.

NMR-activity assays were setup using 500 µM 5'-TCA-oligo with 50 nM A3B$_{CTD}$-AL1 or 2 µM A3B$_{CTD}$-DM in a 50 mM citrate-phosphate buffer with a pH range of 4.5 – 7.5, in 10% D_2O. The ionic strength of buffer was adjusted by the addition of similar levels of sodium phosphate. The initial speed of the reaction was plotted against the indicated pH of the buffer.

4.3.2 Kinetic characterisation of the A3B$_{CTD}$-AL1 and A3B$_{CTD}$-DM activity in deaminating dC of 5'-ATTTCATTT

Assays were setup using the defined optimal conditions (listed above in section 4.3.1) to quantitatively characterise the rate of the deamination of the substrate oligonucleotide by our active A3B$_{CTD}$ variants. ^{1}H NMR spectra were recorded at varying concentrations of the TCA-oligo (50 µM -700 µM) along with 50 nM of A3B$_{CTD}$-AL1 or 2 µM of A3B$_{CTD}$-DM in 50 mM citrate- phosphate buffer (pH 5.5 or 7.5, respectively). Examination of the NMR spectra showed that within an hour (2980 sec, Figure 4.6A) of initiating the reaction an initial conversion of dC to dU occurred, as indicated by the appearance of the H-5 dU doublet peaks (5.73 ppm, Figure 4.6A). After an incubation period of approximately 4

hours ~70 % of the dC was converted to dU (14900 sec, Figure 4.6A). These observations were apparent for both the $A3B_{CTD}$-AL1 and $A3B_{CTD}$-DM enzymes under our given conditions (for simplicity, data shown for $A3B_{CTD}$-AL1, Figure 4.6). To calculate the concentration of conversion of the substrate during the reaction, a doublet of doublets (2.57 to 2.39 ppm) arising from citrate (in the buffer) was used as a concentration standard. The area of the H-5 dC proton doublet (5.92 to 5.88 ppm) was integrated, then was converted into a substrate concentration and plotted against the reaction time. From this the speed of the reaction can be established as it follows a linear relationship (Figure 4.6B). In an ideal situation, it should be possible to obtain K_m and k_{cat} values by direct fitting of the substrate/initial rate data. However, due to the instability of the A3 enzymes during long experiments analysis of product formation becomes complicated. Therefore, the initial rate of deamination was calculated by performing experiments at various substrate concentrations for reliability of K_m values.

Plotting the initial rate of deamination against substrate concentration revealed a hyperbolic plot (see Figure 4.8), consistent with the $A3B_{CTD}$ variant enzymes following Michaelis-Menten kinetics. Initially the reaction rate increases as the substrate increases, but at higher substrate concentrations the reaction rate plateaus as the enzyme is saturated. The data obtained can be fitted with a straight line in a double-reciprocal plot (Lineweaver-Burk plot) by plotting the inverse rate of deamination against the inverse substrate concentration, which is visually easier to analyse. From this plot the Michaelis-Menten constants (K_m, k_{cat}) can be calculated as described in section 2.5.2, and their uncertainties can be determined based on the fit of the experimental data (see Equation 8.1). As each point on this plot represents an individual experiment at various substrate concentrations monitored over time using the same enzyme concentration, more data points were obtained leading to more reliable Michaelis-Menten values than performing the same experiment multiple times. Fitting the data using non-linear least squares using a global fit, we obtain fits and parameter values that are not significantly different from linear analysis (see appendix section Figure 8.4 and Table 8.3)

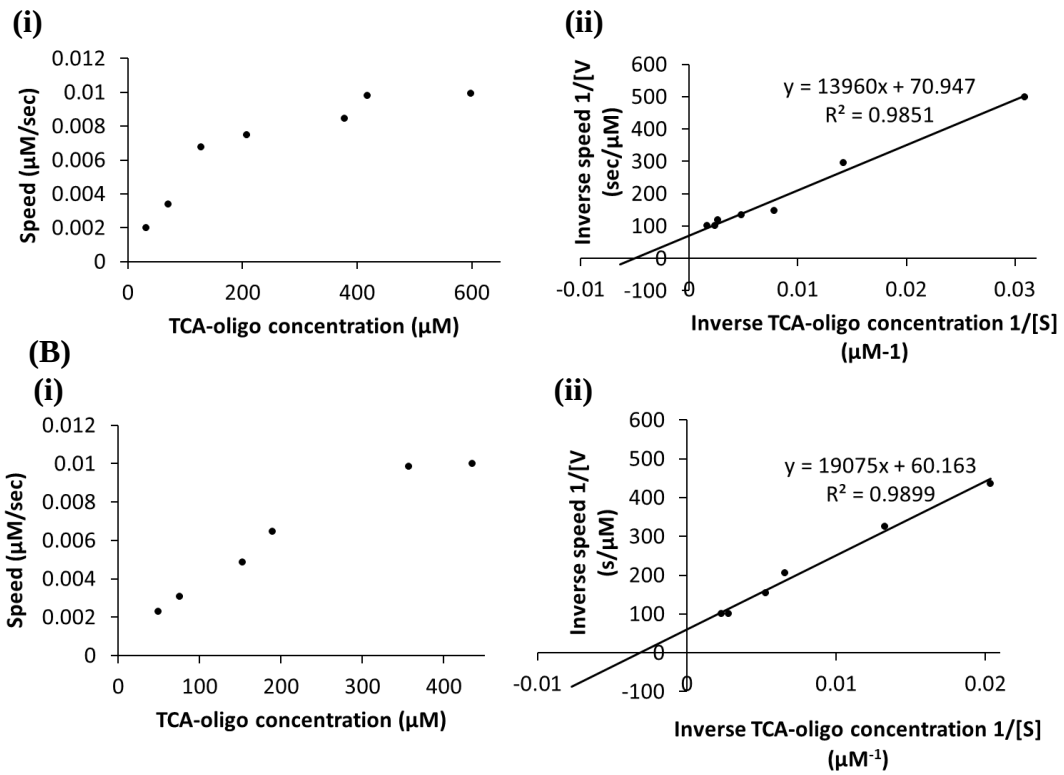


Figure 4.8 Kinetic charachterisation of A3B$_{CTD}$-AL1 and A3B$_{CTD}$-DM cytosine deamination of 5'-ATTTCATTT.

Initial speed of cytosine deamination of the 5'-TCA-oligo at various concentrations (50 – 600 µM) by (A) 50 nM A3B$_{CTD}$-AL1 or (B) 2 µM A3B$_{CTD}$-DM at 298 K. (i) Plot of the initial speed of deamination as a function of substrate concentration, (ii) Lineweaver-Burk plot of the inversed initial speed of deamination as a function of inversed substrate concentration to derive K_m and k_{cat} parameters.

Table 4.1 Kinetic parameters of the active A3B$_{CTD}$ variants determined by ^{1}H-NMR-based assay at 298 K.

Protein	K_m (µM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1} $µM^{-1}$)
A3B$_{CTD}$-AL1swap	197 ± 31	0.28 ± 0.04	0.0014
A3B$_{CTD}$-DM	317 ± 55	0.0082 ± 0.0014	0.000025

Interestingly, the cytosine deamination of the TCA-oligo by our A3B$_{CTD}$-DM was slower than the A3B$_{CTD}$-AL1 variant, reflected by a k_{cat} of 0.0082 s^{-1} for A3B$_{CTD}$-DM, in contrast to a k_{cat} ~ 0.28 s^{-1} for the A3B$_{CTD}$-AL1 variant. These results coincide with A3B$_{CTD}$-DM having a lower affinity for the TCA-oligo in contrast to A3B$_{CTD}$-AL1, as indicated by a higher K_m ~ 320 µM for A3B$_{CTD}$-DM in comparison to a reduced K_m ~ 200 µM for the A3B$_{CTD}$-AL1 variant (see Table 4.1). In addition, the specificity constant defined by

k_{cat}/K_m [127] reinforced that the $A3B_{CTD}$-AL1 variant was ~ 60 fold more efficient at converting the TCA-oligo to TUA-oligo than $A3B_{CTD}$-DM as indicated in Table 4.1, consistent with previous reports [37]. The transfer of the A3A loop 1 into $A3B_{CTD}$ led to a more catalytically active version of $A3B_{CTD}$ [21], attributed to a more open active-site conformation as reported for A3A [23, 36, 239]. Therefore, this variant was used for initial screenings of potential inhibitors discussed further in section 4.4.

4.4 Evaluation of modified-ssDNA for their inhibitory potential on $A3B_{CTD}$ variants using ^{1}H-NMR-based inhibitor assay

A3 inhibitors were designed based on the previously described inhibitors of the closely related CDA enzymes, such as 3-deazauracil (K_i ~ 100 µM) [45], zebularine (K_i ~ 2.3 µM) [6], and 5-fluorozebularine (K_i ~ 0.3 µM) [6]. As A3 enzymes prefer ssDNA substrates [180], these cytosine analogues had to be synthesised as 2'-deoxynucleosides and be incorporated into the preferred A3 ssDNA sequence (referred to as modified-ssDNA). Several modified-oligos including 2'-deoxy-3-deazauridine (3dadU), 2'-deoxy-3-deazazebularine (3dadZ), 2'-deoxyzebularine (dZ), 5-fluoro-2'-deoxyzebularine (displayed in Figure 4.9) were synthesised by Dr. Maksim Kvach and Mr. Harikrishnan Mohana Kurup [135-136] as described in section 2.3.2. For this research, I examined the inhibitory potential of these modified-oligos on the $A3B_{CTD}$ enzymes. As the active site and deamination mechanism for CDA and A3 enzymes are believed to be conserved, we anticipate that a comparable inhibitory effect of the A3 enzymes would arise using these modified oligos.

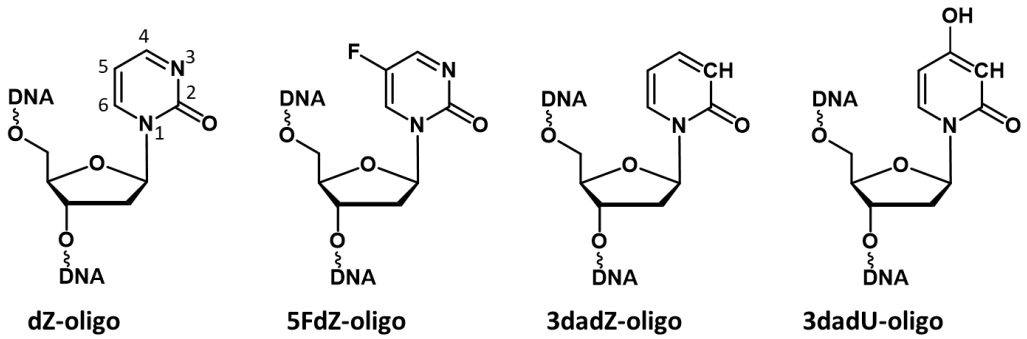


Figure 4.9 Modified oligonucleotides containing cytosine analogues as potential inhibitors of A3. The cytosine of the TCA-oligo is replaced with a modified nucleoside such as 2'-deoxyzebularine (dZ), 5-fluoro-2'-deoxyzebularine (5FdZ), 2'-deoxy-3-deazazebularine (3dadZ), and 2'-deoxy-3-deazauridine (3dadU).

To evaluate our modified oligos, the previously described ^{1}H-NMR-based activity assay was used to directly monitor the rate of deamination of the active $A3B_{CTD}$ on the preferred substrate oligonucleotide in the presence or absence of our modified-oligos. This was followed by the evaluation of the inhibition constant (K_i) [79, 135], as described in section 2.5.3. In a similar manner as described earlier in section 4.3, experiments were conducted at various inhibitor concentrations, which allowed K_i values to be derived from the linear fit of the data in the Dixon plots. Fitting the data using non-linear least squares, we obtain fits and parameter values that are not significantly different from using linear analysis (see appendix section Figure 8.5 and Table 8.4). By using the same experimental conditions, the inhibitory potential of the modified oligos can be evaluated by comparing the K_m value of the substrate to the K_i value of inhibitors.

4.4.1 Qualitative characterisation of the inhibition of cytosine deamination of $A3B_{CTD}$

Our modified-oligos were initially screened for their inhibitory effect on our most catalytically active $A3B_{CTD}$-AL1 variant. Assays were setup using 350 µM of the preferred TCA-oligo along with 50 nM $A3B_{CTD}$-AL1 in the presence of 50 µM modified oligo in the citrate- phosphate pH 5.5 buffer. A series of ^{1}H-NMR spectra were recorded and the initial deamination speed of $A3B_{CTD}$-AL1 in the presence of these modified-oligos were determined (Figure 4.10).

Oligonucleotides possessing the 3dadU and 3dadZ in place of dC led to no change in the initial deamination speed of $A3B_{CTD}$-AL1, relative to the control assay without an inhibitor (Figure 4.10). This indicated that both of these modified oligonucleotides were unable to inhibit $A3B_{CTD}$-AL1 under our conditions. Previous studies reported that the ribose analogue of 3dadU nucleoside is a very weak inhibitor of the human liver CDA with a $K_i \sim 100$ µM [45]. Increasing the concentration of the 3dadU oligo may lead to an observable inhibition of $A3B_{CTD}$-AL1, but as it might only offer a minimal inhibition effect its development was not further advanced. However, the dZ- and 5FdZ-containing oligonucleotides resulted in significant inhibition of $A3B_{CTD}$-AL1, as shown in Figure 4.10 by the substantial decrease in the speed of deamination upon their addition to the reaction. This pinpointed that the dZ-oligo and 5FdZ-oligo are inhibitors of $A3B_{CTD}$.

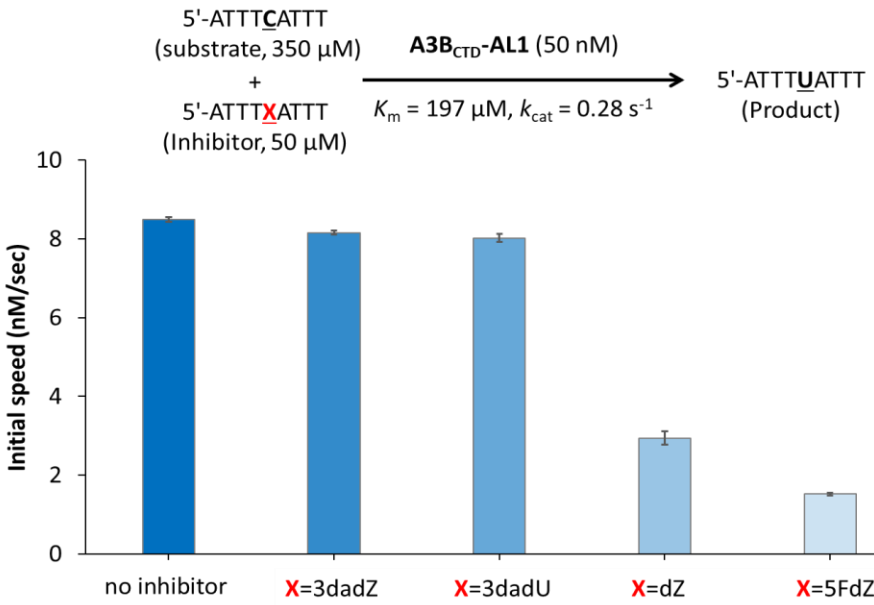


Figure 4.10 Qualitative screen of modified-oligos on the inhibition of A3B$_{CTD}$-AL1 catalysed deamination of the substrate oligonucleotide (5'-ATTTCATTT).

Plot of the initial speed of deamination of 350 µM TCA-oligo by 50 nM A3B$_{CTD}$-AL1 in the presence of 50 µM ssDNA containing the modified nucleoside (X) in citrate-phosphate pH 5.5 buffer at 298 K. Experiments repeated 2-3 times and the mean value was taken and plotted. Error bars reported as standard deviations (SD).

The observed inhibition of the A3B$_{CTD}$ by the 2'-deoxyzebularine derivatives dZ-oligo and 5FdZ-oligo can be described by our inhibitory mechanism model displayed earlier in Figure 3.10 [135-136] (based on the CDA mechanism [281]). As revealed earlier by our thermal shift assay results (refer to section 3.4.2), the catalytic glutamic acid (E255) in the A3's active site is essential for the initial step of protonation of N3 of dZ (and 5FdZ). This protonation makes C4 electrophilic and more susceptible to a nucleophilic attack by a water molecule (or hydroxide ion). Upon this attack the nucleoside (dZ or 5FdZ) transforms into a rigid tetrahedral transition state of cytosine deamination (Figure 3.10). This formation allows dZ (5FdZ) to stay bound in the A3 active-site and thereby inhibit the deamination activity of the enzyme. The lack of inhibition of the A3B$_{CTD}$ enzyme by 3dadZ-oligo showcases the importance of the initial protonation of N3 atom in the inhibitory mechanism, as 3dadZ has a non-protonable CH group in place of an N3 atom. Furthermore, the tetrahedral geometry of C4 in the transition state is equally important for this mechanism, as noted when utilising the 3dadU nucleobase with C4 in planar geometry.

4.4.2 Quantitative characterisation of dZ and 5FdZ-containing ssDNAs as inhibitors of A3B$_{CTD}$

To establish the mode of enzyme inhibition by these inhibitors, the rate of deamination of the substrate oligonucleotide at varying concentrations (50 μM – 700 μM) by A3B$_{CTD}$-AL1 was monitored using the ^{1}H-NMR-based activity assay, with and without the dZ-oligo inhibitor (Figure 4.11). From the Lineweaver-Burk plot, where inverse speed is plotted against inverse substrate concentration, the maximum rate of reaction (V_{max}) is calculated as $1/V_{max}$ at the intercept of the *y*-axis and K_m is determined as $-1/K_m$ from the *x*-axis intercept. This plot revealed that in the absence and presence of the inhibitor the linear regression lines cross the *y*-axis at practically the same position, indicating that V_{max} remains essentially the same (without inhibitor; $1/V_{max}$ = 71 ± 12 s/μM, with inhibitor; $1/V_{max}$ = 83 ± 32 s/μM). Moreover, in the presence of the inhibitor the slope increases leading to an increase in the apparent K_m ~ 450 μM, in comparison to the reaction without inhibitor K_m ~ 200 μM. The decrease in the apparent affinity of the substrate to the enzyme in the presence of the inhibitor is the result of competition of the substrate and inhibitor to bind in the active site of the enzyme. Taken together these results indicate that the modified-oligos are competitive inhibitors of the A3 enzymes. As our inhibitors have similar structures, they are all assumed to be competitive inhibitors.

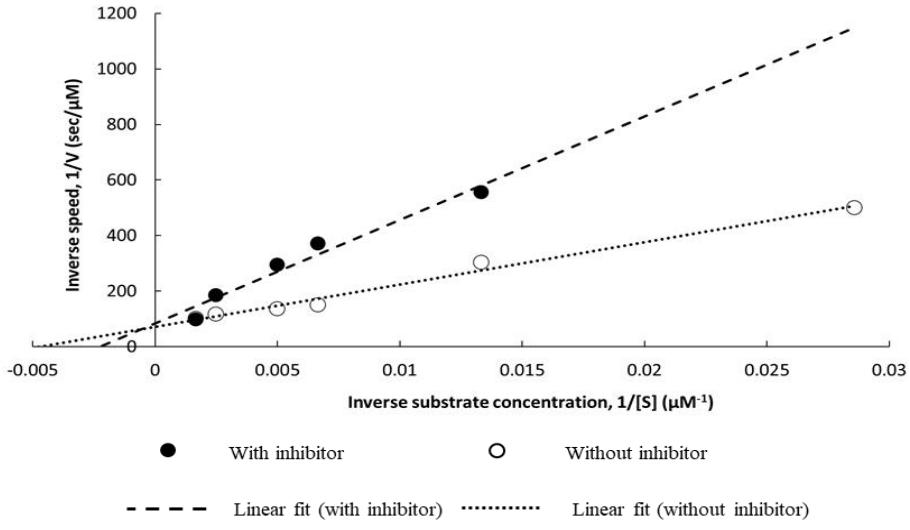


Figure 4.11 Lineweaver-Burk plot of competitive inhibition of A3B$_{CTD}$-AL1 by the dZ-oligo. Deamination rate of the substrate oligonucleotide (5'-ATTTCATTT) by 50 nM A3B$_{CTD}$-AL1 at varied substrate concentrations (50 – 600 μM) in the absence and presence of 20 μM dZ-oligo, in citrate-phosphate pH 5.5 buffer measured at 298 K. Data are plotted as inversed initial speed of deamination as a function of the inversed substrate concentration.

To calculate the inhibition constant (K_i) of the modified oligos (dZ-oligo and 5FdZ-oligo) the deamination reaction of substrate oligonucleotide by $A3B_{CTD}$ was monitored by ^{1}H-NMR (as previously described 4.3.2) in the presence of various inhibitor concentrations. Assuming competitive inhibition for the inhibitors (see earlier Figure 4.11), the linear dependence of the inverse deamination speed as a function of the inhibitor concentration was analysed as described in section 2.5.3. First, we determined that for the dZ-oligo on our $A3B_{CTD}$-AL1 variant K_i was ~7.5 ± 1.7 µM (Figure 4.13B). Comparison of the substrate K_m (~ 200 µM, Table 4.1) to this K_i value, indicated that the dZ-containing oligo had an inhibition potential of K_m/K_i of ~30-fold on the $A3B_{CTD}$-AL1 enzyme. Moreover, the dZ-oligo had essentially the same overall inhibition effect of ~30 fold on the $A3B_{CTD}$-DM variant (K_m/K_i ~320/11; see Table 4.1 and Figure 4.13A). This implied that the replacement of dC with dZ causes the observed inhibitory effect in both $A3B_{CTD}$ variants; therefore, for simplicity we continued our evaluations of the modified-oligos with our most active $A3B_{CTD}$-AL1 variant.

ITC was used as an additional method to compare the binding affinity of the dZ-containing oligo to the $A3B_{CTD}$-AL1 variant. ITC was performed as described in section 2.4.2. The dZ-containing oligo was found to be capable of binding to $A3B_{CTD}$-AL1, under these conditions. This binding event was noted to be driven by a large negative enthalpy (ΔH ~ 42 kcal/mol, Figure 4.12) leading to an overall favourable interaction as indicated by a negative ΔG ~ -7.2 kcal/mol ($\Delta G = \Delta H - T\Delta S$, where T is temperature of 298 K). These results are consistent with earlier observations noting that the presence of a dZ-containing oligo increased the thermal stability of the $A3B_{CTD}$ protein in contrast to the weaker binding affinities of the dC/dU-containing oligos (see Figure 3.11 for the $A3B_{CTD}$-QM-ΔL3 results, and section 3.4.3 for the ITC experiment). Therefore, binding affinities can be determined. We found that under our conditions where 100 µM $A3B_{CTD}$-AL1 was titrated with 300 µM dZ-containing oligo the generated curve reached saturation (Figure 4.12). However, as a consequence of the low protein concentration used to prevent protein aggregation, the curve did not result in the expected sigmoidal shape and lacked an inflection point. Therefore, the binding stoichiometry (N) was fixed to a constant value of 1 (representing a 1 to 1 ratio) to derive the thermodynamic parameters. The directly measured dissociation constant K_d ~ 5.5 ± 0.5 µM obtained from ITC was noted to be comparable to the binding affinity of the dZ-oligo inhibitor (K_i ~ 7.5 ± 1.7 µM, Figure

4.13B). Deriving similar binding affinities using two independent methods further supports that in the dZ-oligo the dZ causes the inhibitory effect.

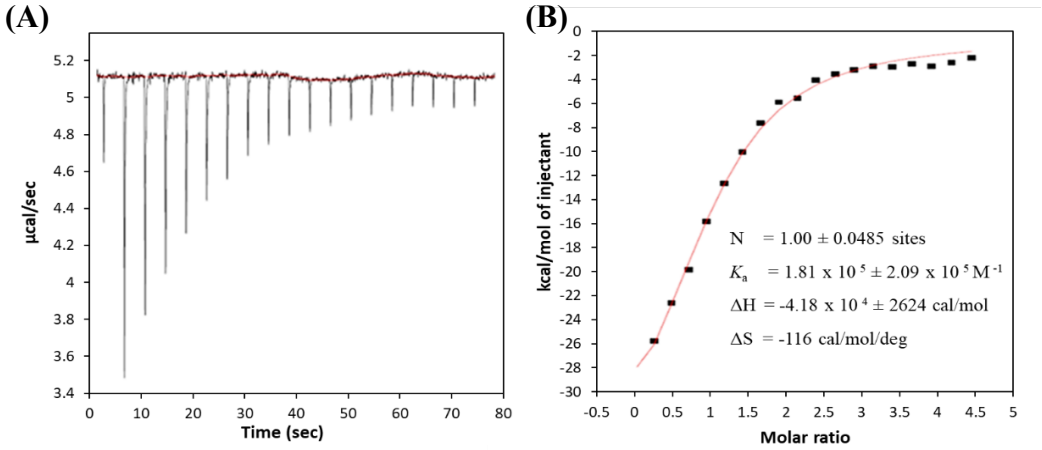


Figure 4.12 ITC experiment of A3B$_{CTD}$-AL1 titrated with dZ-containing oligo.

ITC was performed using 100 µM A3B$_{CTD}$-AL1 titrated with 300 µM dZ-containing oligo (5'-ATTTdZATTT) in 50 mM citrate-phosphate pH 5.5 buffer at 25 °C as described in section 2.4.2. (A) Raw injection data, (B) data of integrated injection enthalpies fitted with a one-site binding model to obtain the thermodynamic parameters; enthalpy (ΔH), entropy (ΔS), equilibrium binding association constant (K_a), and stoichiometry (N).

Next, we determined the inhibition potency of the 5FdZ-oligo on our A3B$_{CTD}$-AL1 variant. We found that the 5FdZ-oligo (K_i ~ 2.1 ± 0.8 µM) had a 3.5 times lower K_i than the dZ-oligo (K_i ~ 7.5 ± 1.7 µM) (see Figure 4.13B and C), indicating that the 5FdZ-containing oligo was much more potent than the dZ-containing oligo (dZ- ~ 30 fold, 5FdZ ~ 100 fold). These findings were also echoed by results obtained by our collaborators (Assoc. Prof. Daniel Harki's team, University of Minnesota) on an active wild-type A3A (expressed in human cells), which was not surprising as the residues in the active-site of A3B$_{CTD}$ and A3A are conserved [239].They used a fluorescence polarisation competition assay, whereby a fluorescently labelled ssDNA containing a TC-motif was pre-bound to A3A then was competed with unlabelled dZ- and 5FdZ-oligo inhibitors to assess their binding affinities and calculate the half maximal inhibitory concentration (IC$_{50}$). These assays demonstrated that the 5FdZ-containing oligo (IC$_{50}$ ~ 0.16 ± 0.01 µM) was ~2.4 times a more potent inhibitor of the wild-type A3A than the dZ-containing oligo (IC$_{50}$ ~ 0.39 ± 0.03 µM) [136], in a similar manner as determined for our A3B$_{CTD}$-AL1.

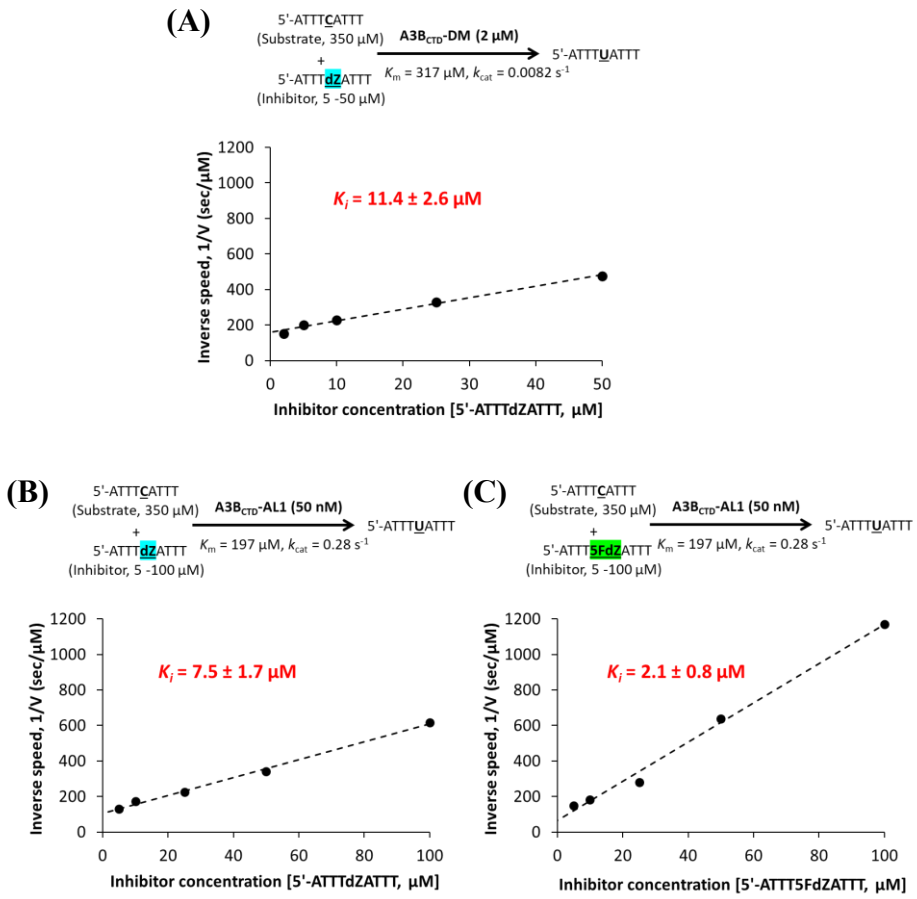


Figure 4.13 Dixon plots of the inhibition of the A3B$_{CTD}$ variants by dZ- and 5FdZ-containing oligos monitored by ^{1}H-NMR inhibition assay.

Deamination of the 350 µM TCA-oligo (5'-ATTTCATTT) by (A) 2 µM A3B$_{CTD}$-DM in the presence of varying concentration (2-50 µM) of the dZ-containing oligo citrate-phosphate pH 7.5 buffer, (B) 50 nM A3B$_{CTD}$-AL1 in the presence of varying concentration (5-100 µM) of the dZ-containing oligo citrate-phosphate pH 5.5 buffer, (C) 50 nM A3B$_{CTD}$-AL1 in the presence of varying concentration (5-100 µM) of the 5FdZ-containing oligo citrate-phosphate pH 5.5 buffer. Measured by ^{1}H-NMR-based inhibition assay at 298 K .

It is thought that due to the electron-withdrawing fluorine of 5FdZ in the heterocycle [136] causes C4 to be become more electrophilic and therefore more susceptible to the nucleophilic attack of $H_2O/^-OH$. This enhances inhibition of A3B$_{CTD}$ and A3A by the 5FdZ-oligo in comparison to the dZ-oligo. This observed trend of inhibition was consistent with previous reports where the 5-fluorozebularine was determined to be a better inhibitor of CDA than zebularine and 3-deazauracil (K_i ~ 0.3 µM, 2.3 µM, [6] and 100 µM [45], respectively). This observation further reinforces our rationale that the structure of the cytosine analogues, placed instead of the dC in the preferred A3 ssDNA substrate, primarily determines the inhibitory potential of the modified oligonucleotides on the A3 enzymes.

4.5 Conclusions

Two active $A3B_{CTD}$ variants, $A3B_{CTD}$-AL1 and $A3B_{CTD}$-DM, were purified to study their catalytic deamination activity on the preferred substrate oligonucleotide (TCA-oligo) and to evaluate this activity in the presence of modified oligonucleotides. Using these active $A3B_{CTD}$ variants we characterised their kinetic parameters in regard to their deamination activity on the dC in the preferred TCA-oligo, using the previously described ^{1}H-NMR based activity assay [94]. We found that the $A3B_{CTD}$-AL1 variant (K_m ~ 200 µM) was significantly more active than the $A3B_{CTD}$-DM (K_m ~ 320 µM) variant, suggestive that the replacement of loop 1 of $A3B_{CTD}$ with that from A3A enhanced the activity of our $A3B_{CTD}$-AL1 variant by mimicking a more open active-site conformation allowing easier binding of ssDNA, as shown in the A3A structure, Figure 1.15A [23, 36, 239].

Using $A3B_{CTD}$-AL1 variant (as it was the most active $A3B_{CTD}$) our modified oligonucleotides containing cytosine analogues 3dadU, 3dadZ, dZ, and 5FdZ were evaluated for their inhibitory potential. We discovered that the dZ- and 5FdZ-containing oligos significantly inhibited the deamination activity of $A3B_{CTD}$-AL1, whereas 3dadZ- and 3dadU-containing oligos had no impact on the activity. These results validated our proposed inhibition mechanism (described in Figure 3.10), which highlights the importance of the protonation of N3 in dZ by the A3's active-site catalytic glutamic acid (E255), which triggers the conversion of the dZ nucleoside into the rigid hydrated tetrahedral transition state of cytosine deamination. Therefore, protonation of the N3 of dZ and the 5FdZ nucleosides initiates this reaction allowing their transition states to remain tightly bound in the A3's active site and thereby blocking further catalysis by the enzyme. On the other hand, for 3dadZ, which contained a CH group in place of the N3 atom and for 3dadU, for which C4 had a planar geometry in the transition state, the lack of inhibition of both these nucleoside-containing oligos demonstrated the importance of the protonation of the N3 atom and the tetrahedral geometry of C4 in the transition state to achieve inhibition.

Using our ^{1}H-NMR based inhibitor assay, the inhibition constants of the dZ-oligo on the activity of the $A3B_{CTD}$ variants were characterised first. The dZ-oligo was determined to be binding competitively into the active-site of $A3B_{CTD}$ and as our cytidine analogue inhibitor (5FdZ) had a similar structure we presumed that it would also be competitive. The dZ-containing oligo was found to be capable of inhibiting both our $A3B_{CTD}$ variants with comparable low micromolar inhibition constants ($A3B_{CTD}$-AL1 K_i ~ 7.5 µM,

A3B$_{CTD}$-DM K_i ~ 11 μM). The overall inhibition effect (K_m/K_i ~ 30 fold) of this oligonucleotide on the A3B$_{CTD}$-DM and A3B$_{CTD}$-AL1 variants were the same. Furthermore, the 5FdZ-containing oligo was determined to be a more potent inhibitor than the dZ-containing oligo for both our A3B$_{CTD}$-AL1 (K_m/K_i ~ 100 fold) and A3A (as reported by our collaborators [136]), in a similar manner as the previously reported 5-fluorozebularine on CDA [6]. Moreover, Dr. Stefan Harjes also showed that the incorporation of dZ or 5FdZ into the A3G's preferred deamination CC<u>C</u>-motif (underlined is preferred dC replaced with dZ or 5FdZ) in short ssDNA (~10 nucleotides) inhibited both A3G$_{CTD}$ and a full-length A3G (expressed in human cells) with low micromolar inhibition constants (further elaborated on in section 6.4.3.1 and Figure 6.21) [135-136]. Thus, the structure of the cytosine analogue nucleoside must directly influence the inhibition of the A3 enzymes. Notably, inhibition of A3G$_{CTD}$ was comparable to the inhibition of the full-length A3G [18], indicating that the catalytic activity is largely attributed to the CTD in two-domain A3 enzymes. This result justifies the rationale to investigate modified-ssDNA inhibitors on A3B$_{CTD}$ to illustrate inhibition of the full-length A3B.

Overall, we discovered the first specific substrate-like ssDNA inhibitors capable of targeting A3 enzymes. These nucleoside-containing oligonucleotides would not target CDA, as CDA prefers free nucleosides over oligonucleotides as reported in the literature [14, 120]. Moreover, as most A3 enzymes (except A3G) prefer to deaminate a dC within a 5'-TC-motif, replacement of this dC with the cytosine analogue inhibitors (dZ or 5FdZ) would specifically target most of these A3 enzymes, as indicated by the inhibition of both A3B$_{CTD}$ and A3A. As A3B has been noted to drive the mutagenic evolution of cancer cells, its selective inhibition is crucial, whilst preserving the action of the other A3 enzymes involved in the innate immunity (discussed further in chapter 5).

5. Selective inhibition of A3 enzymes by substrate-like inhibitors containing 2’-deoxyzebularine

5.1 Introduction

The majority of the work reported in this chapter was incorporated from our published journal article:

1) Barzak, F. M., Harjes, S., Kvach, M. V., Kurup, H. M., Jameson, G. B., Filichev, V. V., & Harjes, E. Selective inhibition of APOBEC3 enzymes by single-stranded DNAs containing 2′-deoxyzebularine. *Organic & Biomolecular Chemistry*. 2019, 17 (43), 9435-9441. Doi: 10.1039/c9ob01781j [8].

   In this article, I hypothesised and tested whether selective inhibition using the differentially deaminated suboptimal CCC-oligo was feasible. First, I expressed and purified the $A3B_{CTD}$ variant enzymes ($A3B_{CTD}$-QM-ΔL3-AL1swap (termed here A3A-mimic) and $A3B_{CTD}$-DM), then performed ^{1}H-NMR based activity assays to characterise the kinetic deamination activity of these enzymes on the preferred dC in the CCC-oligo, dZCC-oligo, and CCdZ-oligo. I characterised the inhibition potency of dZ-containing oligonucleotides using the ^{1}H-NMR based inhibition assay. I performed all analysis of the derived data and wrote the first draft of article.

The work reported in this chapter was also presented in the conference listed below.

> Barzak, F. M., Kvach, M. V, Harjes S., Jameson G.B, Aihara, H., Harris R. S., Harki, D. A., Filichev, V. V., Harjes, E. DNA-based inhibitors of the human APOBEC3B DNA cytosine deaminase. Selected oral and poster presentations at the NZMS/NZSBMB 'Microbes and molecules' Meeting 2018 held between the 26th – 29th November 2018 at the University of Otago, Dunedin, New Zealand.

A central innate immune defence mechanism against pathogens occurs through the deamination of cytosine in foreign ssDNA by the A3 enzymes [99]. In addition, human nuclear and mitochondrial genomes are susceptible to editing of their cytosines by several A3 enzymes, predominantly A3A, A3B, and A3G [255]. In particular, A3B has been identified as the major source of genetic mutations in multiple cancers, which in turn leads to the development of drug resistance. [34, 82, 146, 244, 268, 270, 295]. Due to the fundamental role of A3 enzymes in the defence against pathogens, the selective inhibition of A3B is vital to suppress its mutagenic action, while maintaining the activity of other A3 enzymes.

Selective small-molecule A3G inhibitors have been reported [151, 196]. We described the first substrate-like competitive A3 inhibitor with a low micromolar inhibition constant (refer to section 4.4) [135], in which the target dC was substituted with dZ in the preferred ssDNA (5'-ATTTCATTT) of our $A3B_{CTD}$ variant. However, as most A3 enzymes (except A3G) (see Table 1.2) have the same intrinsic preference of deaminating a cytosine

preceded by a thymine, this inhibitor is not selective and it is uncertain as to whether a selective A3B inhibitor can be obtained. In this chapter, we first demonstrate that neighbouring nucleotides in the ssDNA sequence influence the deamination selectivity of our engineered A3 enzymes in the same way as they do for the wild-type sequences (see section 4.3). Utilising this knowledge, we designed and tested oligonucleotides containing dZ surrounded by different nucleotides. This illustrated that the nucleobases in the oligonucleotide strongly influence the specificity of our inhibitor, leading to the first selective oligonucleotide inhibitor of A3 enzymes. These findings provide a platform for further development of selective A3 inhibitors.

5.2 Preference of A3 enzymes to deaminate cytosine is dependent on neighbouring nucleotides in sequence

Although A3A and A3B favour the deamination of a T<u>C</u> motif (underlined C is deamination target) where the target dC is preceded by a thymidine, these enzymes are also capable of deaminating cytosines in ssDNA with other adjacent nucleobases, although with lower efficiency [36, 252]. One example, wild-type A3A and $A3B_{CTD}$ can deaminate dC within sequences with a CCC-motif [37], which is the most favoured deamination motif for the A3G ($A3G_{CTD}$, 5'-ATTCC<u>C</u>AATT, K_m ~ 570 μM [94]). A3A and A3G preferentially deaminate dC at the 3'-end of the CC<u>C</u>-motif, where the target dC is followed by a purine (dA or dG) at the 3'-end [37, 41, 122-123, 207, 252]. A3B prefers the dC at the 5'-end of this motif, denoted <u>C</u>CC [36-37], flanked by a thymidine in the adjacent 5'-position. The other non-preferred dCs in the CCC motif are less efficiently deaminated by A3A and A3B as the recognition site is not favoured by the enzymes [37]. Moreover, $A3G_{CTD}$ does not deaminate dC at the 5' position of the <u>C</u>CC-motif and has little enthusiasm for the middle dC of the C<u>C</u>C-motif (5'-ATTC<u>C</u>UAATT K_m ~ 3.6 mM) [94]. Thereby, nucleobases surrounding the target dC significantly influence the substrate recognition of the A3 enzymes [37, 47, 208].

5.2.1 Evaluation of the deamination preference of cytosines within the 5'-ATTCCCAATT sequence by our active $A3B_{CTD}$ variants

Based on the literature, we wanted to examine how our active A3 variants, $A3B_{CTD}$-AL1 and $A3B_{CTD}$-DM (described in chapter 4), deaminate dC in a suboptimal substrate oligonucleotide containing the CCC-motif. Several oligonucleotides were used in this study with the rationale described in Table 5.1.

Table 5.1 Oligonucleotides used in this study

Name	DNA sequence	Rational
TCA-oligo	5'-ATTTCATTT	Used as preferred substrate sequence of A3A and $A3B_{CTD}$ [37]
CCC-oligo	5'-ATTCCCAATT	The 3rd dC (from 5' to 3' ends) is the preferred substrate of $A3G_{CTD}$-WT [94] and preferentially deaminated by A3A [37]. The 1st dC is preferentially deaminated by $A3B_{CTD}$ [37]
UCC-oligo	5'-ATTUCCAATT	Used as a standard for the chemical shift of the H-5 proton doublet of the 1st dU, as a partial product in NMR assay
CUC-oligo	5'-ATTCUCAATT	Used as a standard for the chemical shift of the H-5 proton doublet of the 2nd dU, as a partial product in NMR assay
CCU-oligo	5'-ATTCCUAATT	Used as a standard for the chemical shift of the H-5 proton doublet of the 3rd dU, as a partial product in NMR assay
CUU-oligo	5'-ATTCUUAATT	Used as a standard for the chemical shift of the H-5 proton doublet of the 2nd and 3rd dU, as a partial product in NMR assay
UUU-oligo	5'-ATTUUUAATT	Used as a standard for the chemical shift of the H-5 proton doublet of the 1st – 3rd dU, as a full product in NMR assay

First, 1D ^{1}H-NMR spectra were recorded of the substrate CCC-oligo (5'-ATTCCCAATT) and subsequent standard oligonucleotides containing dU in various positions (described in Table 5.1, partial and full products), to confirm that the chemical shifts of the H-5 proton doublet of dU can be used for monitoring of product formation. The H5-proton doublet of dU in the 5' position in of the UCC-oligo appeared at 5.775 ppm (J = 8.50 Hz), a doublet for dU in the middle position of CUC appeared at 5.748 ppm (J = 8.43 Hz), and the doublet for dU at the 3' position in CCU appeared at 5.67 ppm (J = 8.49 Hz). The H-5 signals of uracils in these oligonucleotides were well dispersed (Figure 5.1), simplifying future detection of product (dU) peaks as a result of A3-catalysed deamination of the individual cytosines in the CCC-oligo using ^{1}H NMR [79, 94].

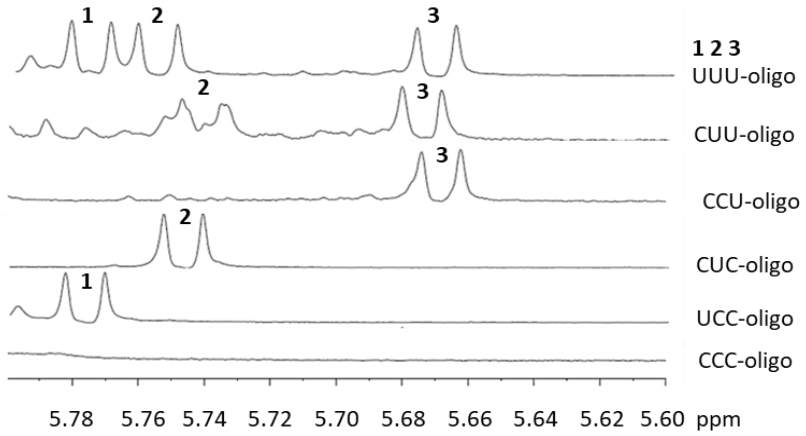


Figure 5.1 1D ^{1}H-NMR spectra of CCC-oligo and standards of possible products of dC deamination.

1H-NMR spectra recorded on a 700-MHz spectrometer using 500 µM oligonucleotide in 50 mM citrate-phosphate pH 5.5 buffer (50 mM citrate-phosphate, 200 mM NaCl, 2 mM β-ME, 200 µM DSS, pH 5.5) containing 10 % D_2O. Peaks are numbered to indicate the position of dU in the oligonucleotide sequence.

Next, we monitored the deamination pattern of the CCC-oligo by $A3B_{CTD}$-AL1 and $A3B_{CTD}$-DM. Experiments were setup with an excess of 800 µM CCC-oligo along with 2 µM of either A3 variant in 50 mM citrate-phosphate buffer at their respective pH optimum ($A3B_{CTD}$-AL1 in pH 5.5 buffer, $A3B_{CTD}$-DM in pH 7.5 buffer, Figure 4.7) and the deamination rates were monitored using ^{1}H-NMR. Our $A3B_{CTD}$-DM variant favoured the deamination of the first dC at the 5' position of the CCC motif (Figure 5.2B), consistent with the selective deamination behaviour reported for the wild-type $A3B_{CTD}$ [37]. In contrast, our $A3B_{CTD}$-AL1 variant favoured the deamination of the third cytosine in the CCC motif (dU position 3, Figure 5.2A), similar to the preference of the wild-type A3A and $A3G_{CTD}$.

The observed variation in the deamination preference of the CCC-oligo between these $A3B_{CTD}$ variants (Figure 5.2) was caused by the transplant of loop 1 of A3A into $A3B_{CTD}$, leading to the transfer of A3A characteristics to $A3B_{CTD}$, which was previously reported [37]. This loop swap not only affected CCC-motif substrate specificity, but as described earlier led to a more catalytically active $A3B_{CTD}$ construct in comparison to A3BCTD-DM (see Table 4.1) and shifted the pH optimum from pH 7.5 to 5.5 (Figure 4.7) as discussed in section 4.3. Therefore, in regard to the deamination of the CCC-oligo, the $A3B_{CTD}$-AL1 variant reflects the selective deamination of the wild-type A3A in ssDNA substrates containing several cytosines [37]. Thus, the $A3B_{CTD}$-AL1 variant can be used to

model A3A in the selective deamination of the CCC-oligo and will therefore be referred to as A3A-mimic in the remainder of this chapter.

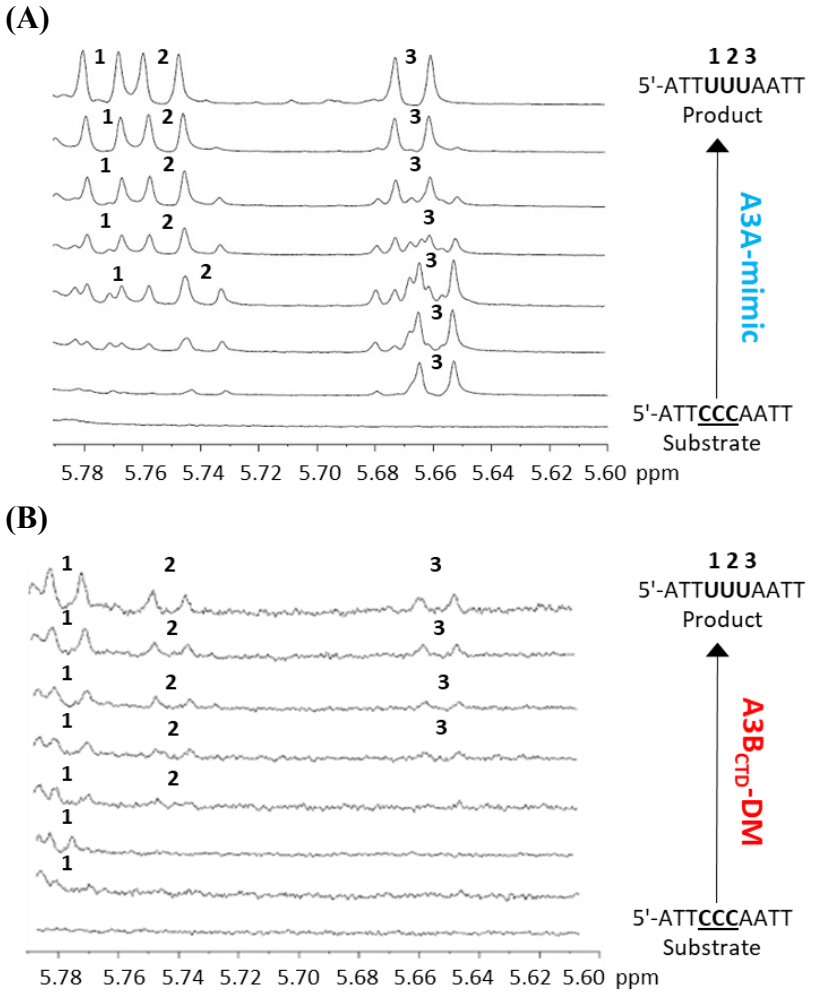


Figure 5.2 1D ^{1}H-NMR spectra of the deamination of the CCC-oligo by our active A3 variants. Assays were setup using 800 µM CCC-oligo with either (A) 2 µM A3B$_{CTD}$-AL1 (A3A-mimic) or (B) 2 µM A3B$_{CTD}$-DM, in 50 mM citrate-phosphate buffer containing 200 mM NaCl, 2 mM β-ME, 200 µM DSS, pH 5.5 or pH 7.5, respectively, in 10 % D_2O. The experiment was monitored using ^{1}H-NMR recorded at 298 K. Note that the spectra shown here are recorded every 15 minutes and that the substrate has no signal in this region at the beginning of the reaction.

Regarding Figure 5.2B, we observed that A3B$_{CTD}$-DM systematically deaminates the dC in position 1 of the CCC motif first, followed by the middle dC (UCC) and finally the dC at position 3 (UUC), under our conditions. On the other hand, A3A-mimic first deaminates the dC in position 3 (CCU), but subsequent deamination of the CCU species is random forming multiple intermediate species until all the dC nucleotides are deaminated, as displayed in Figure 5.2A. The deamination of only one dC at the beginning of the reaction (Figure 5.2) indicates that the enzyme dissociates from the oligonucleotide

between deamination events, since the concentration of the substrate greatly exceeds the concentration of the enzyme. In addition, the deamination products of the less preferred dCs were detected at later time-points and were always less intense in comparison with the signal from the preferred dC indicating that deamination of the non-preferred cytosines was incomplete during the time frame of this experiment (~2 hours, Figure 5.2).

5.2.2 Kinetic characterisation of A3A-mimic and A3B$_{CTD}$-DM for deamination of preferred dC in 5'-ATTCCCAATT oligonucleotide

To evaluate the rate of deamination of the A3 variants on their preferred dC within the CCC-oligo, the aforementioned NMR-based activity assay was performed [94] (as described in sections 2.5.2 and 4.3). Typically, direct fitting of the rate of substrate conversion is possible to obtain the K_m and k_{cat} values. However, for A3 enzymes several factors complicate the analysis of product formation over long periods of time. First, during long experiments the enzymes tend to become unstable; second, when substrate concentration decreases throughout the experiment the A3 enzymes begin to deaminate other cytosines in the CCU- and UCC-motifs. Therefore, to address these issues the initial deamination rate of the preferred cytosine (first dC deaminated) was obtained using NMR assays.

A series of ^{1}H NMR spectra was recorded of the CCC-oligo at varying concentrations (50 – 750 µM) with either 50 nM A3A-mimic or 2 µM A3B$_{CTD}$-DM in their respective buffers (citrate-phosphate buffer at pH 5.5 or pH 7.5). The initial speed of the reaction was calculated as described in Materials and Methods 2.5.2 (using the dU H-5 proton doublet of CC<u>U</u> at 5.67 ppm and the H-5 proton doublet of <u>U</u>CC at 5.775 ppm, see Figure 5.1). Plotting the initial speed of deamination against substrate concentration a hyperbolic plot is obtained, where the A3 enzymes become saturated at high substrate concentrations, indicating Michaelis-Menten kinetics (Figure 5.3A i and Figure 5.3B i). From this the Lineweaver-Burk plot (1/rate versus 1/[substrate]) was fitted with linear regression (Figure 5.3A ii & Figure 5.3B ii) and the Michaelis-Menten constants were derived as described in section 2.5.2. Fitting the data using non-linear least squares using a global

fit, we obtain parameter values that are not significantly different from the linear analysis (see appendix Figure 8.6 and Table 8.3)

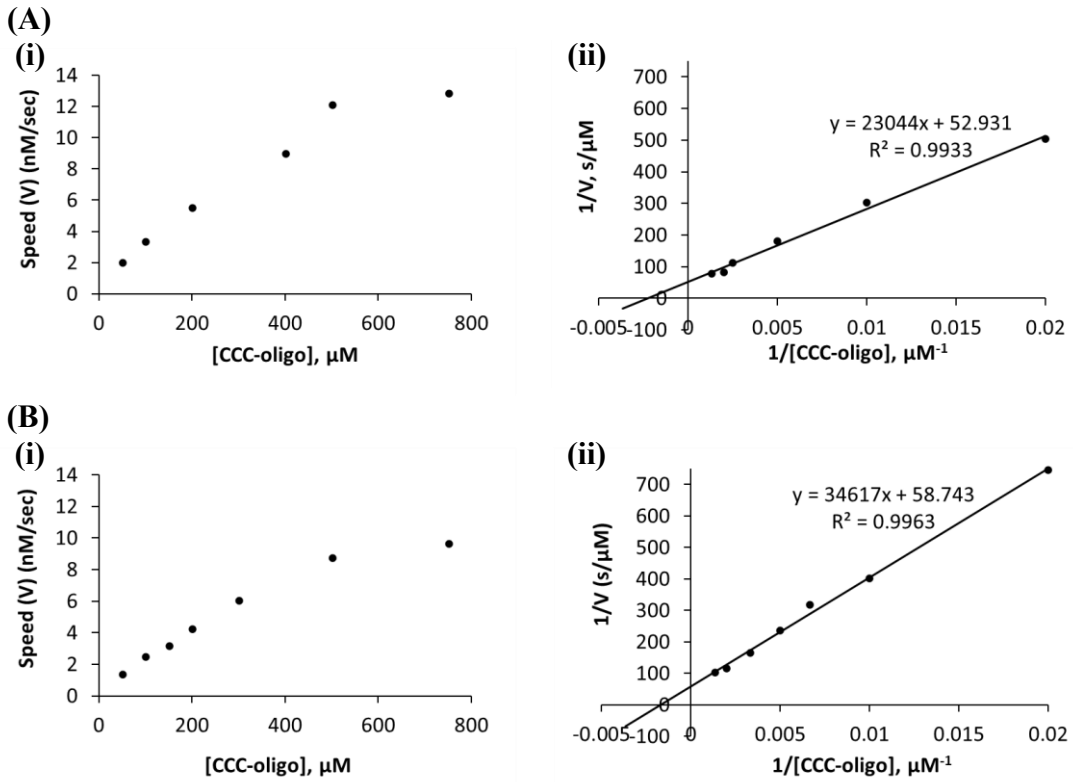


Figure 5.3 Kinetic characterisation of A3A-mimic and A3B$_{CTD}$-DM on the deamination of their preferred dC within 5'-ATTCCCAATT.

(A) Speed of deamination of the dC at the 3'-end in the CC<u>C</u>-oligo with 50 nM A3A-mimic (A3B$_{CTD}$-AL1) in 50 mM citrate-phosphate pH 5.5 buffer, (B) speed of deamination of the dC near the 5'-end in the CC<u>C</u>-oligo with 2 µM A3B$_{CTD}$-DM in 50 mM citrate-phosphate pH 5.5 buffer. Speed of deamination was obtained using the ^{1}H-NMR based-activity assay at 298 K with varying substrate concentrations (20 – 800 µM) (i) Plot of the initial speed of deamination as a function of substrate concentration, (ii) Lineweaver-Burk plot of the inversed initial speed of deamination as a function of inversed substrate concentration to derive K_m and k_{cat} parameters.

The apparent binding affinities of our A3 variants on their preferred dC within the CCC-oligo (A3A-mimic K_m ~ 400 µM for CC<u>C</u>, A3B$_{CTD}$-DM K_m ~ 600 µM for <u>C</u>CC) were determined to be 2-fold weaker than the target dC in the T<u>C</u>A-oligo (where for A3A-mimic (A3B$_{CTD}$-AL1) K_m ~ 200 µM, and for A3B$_{CTD}$-DM K_m ~ 300 µM, as described in Table 4.1) [135]. Furthermore, the deamination of the TCA-oligo by the A3 enzyme was 60 times more efficient than that of the enzymes' preferred dC in the CCC-oligo, as implied by the specificity constant (k_{cat}/K_m [127], Table 5.2), consistent with the CCC-motif being a suboptimal substrate for A3A and A3B$_{CTD}$ [37].

Table 5.2 Kinetic parameters of A3A-mimic and A3B$_{CTD}$-DM determined by the ^{1}H-NMR-based activity assay at 298 K.

Enzyme	Oligonucleotide	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1} μM^{-1})
A3A-mimic	CC**C**-oligo	440 ± 80	0.38 ± 0.06	0.00086
	T**C**A-oligo [135]	197 ± 30	0.28 ± 0.041	0.0014
A3B$_{CTD}$-DM	**C**CC-oligo	590 ± 90	0.009 ± 0.001	0.000015
	T**C**A-oligo [135]	317 ± 55	0.0082 ± 0.0014	0.000025

5.3 Evaluation of dZCC and CCdZ oligonucleotides as selective inhibitors of A3

The first modified-oligo inhibitor of the A3s was developed by incorporating dZ into the enzymes' preferred deamination motif (5'-TC), as discussed earlier in section 4.4 [135]. The dZ-containing oligo was non-selective as it inhibited both A3A-mimic (A3B$_{CTD}$-AL1) and A3B$_{CTD}$-DM with similar inhibition constants (discussed in section 4.4.2), as both A3A and A3B preferred deaminating cytosines preceded by thymine [135]. However, based on the nature of selective deamination of A3A and A3B$_{CTD}$ on DNA substrates containing several cytosines, selective inhibition of A3B was theorised to be feasible. We hypothesised that by replacing the dC at the 5' end of the CCC motif by dZ, the 5'-dZCC oligo would selectively inhibit A3B, and conversely by placing dZ near the 3'-end of the CCC motif, the 5'-CCdZ oligo would inhibit A3A and A3G, but not A3B.

To test this hypothesis, modified-oligos with dZ incorporated into the CCC-oligo as CCdZ and dZCC oligos (rationale described in Table 5.3) were synthesised by Dr Maksim Kvach and Harikrishnan Mohana Kurup as described in section 2.3.2 [8]. Placement of dZ in either the middle position (CdZC-oligo, dU labelled 2 Figure 5.1) or in both the 5' and 3' position (dZCdZ-oligo, dU labelled 1 and 3 Figure 5.1) of the CCC-oligo was expected to lead to non-selective inhibition of the A3 variants, so was not explored. To investigate selective inhibition of the A3 enzymes by the above-mentioned modified oligos, the A3B$_{CTD}$-DM variant and A3A-mimic were employed to model A3B$_{CTD}$ and A3A, respectively. The TCA-oligo was a better substrate for the A3A-mimic and A3B$_{CTD}$-DM than the CCC-oligo (CCC-oligo K_m ~2 fold higher than TCA-oligo, see Table 5.2), therefore the TCA-oligo was used as a substrate in our NMR-based inhibition assay to evaluate these modified-oligonucleotides.

Table 5.3 Modified oligonucleotides used in this study

Name	DNA sequence	Rational
dZ-oligo	ATTTdZATTT	Non-selective inhibitor of A3A and $A3B_{CTD}$ [135]
CC**dZ**-oligo	ATTCCdZAATT	Modified oligonucleotide to compare with the substrate (CCC) and study inhibition of A3A-mimic. Determine if it is a substrate of $A3B_{CTD}$ (the 1st dC). Inhibitor of $A3G_{CTD}$-WT [94]
dZCC-oligo	ATTdZCCAATT	Modified oligonucleotide to compare with the substrate (CCC) and study inhibition of $A3B_{CTD}$. Determine if it is a substrate of A3A and $A3G_{CTD}$ (the 2nd dC in the sequence)

5.3.1 Qualitative characterisation of the inhibition of A3A-mimic and $A3B_{CTD}$-DM by the dZCC and CCdZ-oligos

To assess whether selective inhibition of our A3 variants can be achieved using the CCdZ- and dZCC-oligos, an initial screen was performed. The deamination of 350 µM substrate TCA-oligo by our A3 variants was monitored by ^{1}H-NMR in the absence and presence of 50 µM oligonucleotides (CCC-oligo, CCdZ-oligo, or dZCC-oligo), to examine the residual enzyme activity (see Figure 5.4 and Figure 5.5).

Figure 5.4 shows the setup of the assay using 350 µM TCA-oligo deaminated by 50 nM A3A-mimic in the absence and presence of 50 µM CCdZ-oligo monitored by ^{1}H-NMR under our standard conditions (for simplicity data shown for $A3B_{CTD}$-AL1, Figure 5.4). The H-5 proton doublet signal of the cytosine from the TCA-oligo (Figure 5.4A i and B i) was integrated then converted into a concentration and plotted against the time of the reaction to determine the speed of the reaction from the linear regression equation (demonstrated in Figure 5.4A ii and B ii). Examination of the NMR spectrum showed that the presence of the CCdZ-oligo significantly slows the A3A-mimic's rate of deamination (Figure 5.4B ii) as observed by the lack of dC to dU conversion (5.73 ppm, Figure 5.4B i), in comparison to the control experiment without the CCdZ-oligo (Figure 5.4A). Consistent with our hypothesis, the CCdZ-oligo was found to significantly inhibit A3A-mimic (Figure 5.5A) (also the $A3G_{CTD}$ as reported in [135]), but not $A3B_{CTD}$-DM (CCdZ, Figure 5.5B). In contrast, dZCC-oligo inhibited the $A3B_{CTD}$-DM variant (Figure 5.5B) but no inhibition of A3A-mimic (dZCC, Figure 5.5A) (or $A3G_{CTD}$ [135]) was detected.

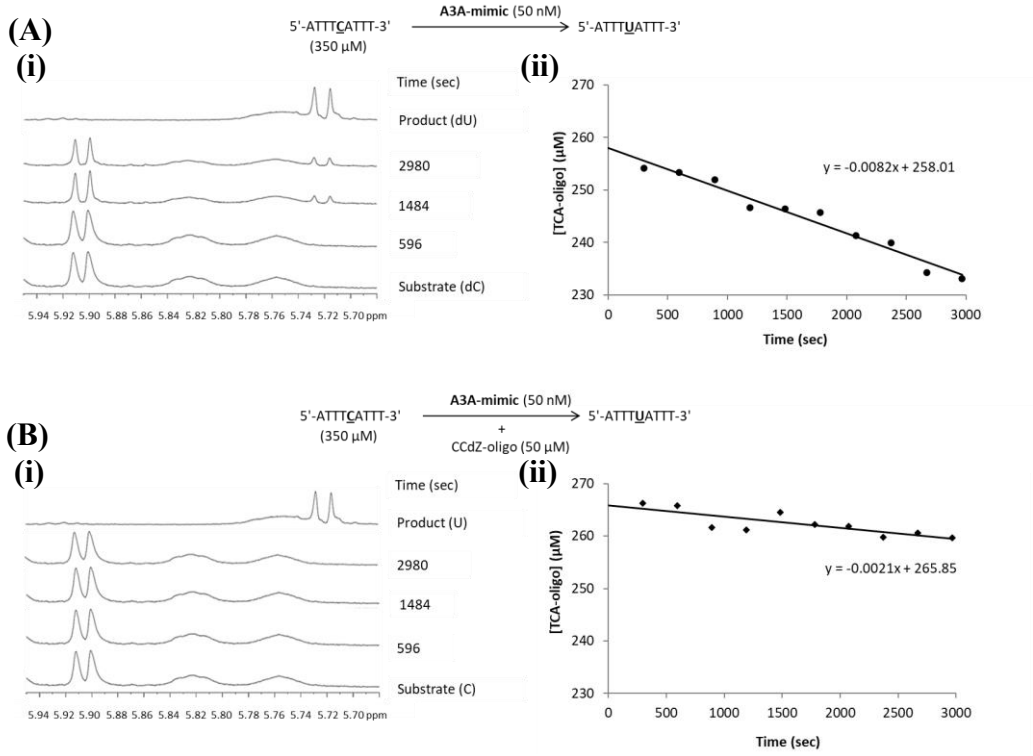


Figure 5.4 Time resolved deamination of TCA-oligo by A3A-mimic in the absence and presence of the CCdZ-oligo and assessment of the initial speed of the reaction using ^{1}H-NMR.

A series of ^{1}H-NMR spectra was recorded on a 700-MHz spectrometer every 298 sec using 350 µM TCA-oligo (substrate) during the deamination reaction by 50 nM A3A-mimic ($A3B_{CTD}$-AL1) in citrate-phosphate pH 5.5 buffer, with 10% D_2O in the (A) absence or (B) presence of 50 µM CCdZ-oligo at 298 K. (i) ^{1}H-NMR spectra at various time points during the deamination reaction. (ii) Plot of substrate concentration (TCA-oligo) versus time of the reaction fitted with linear regression equation.

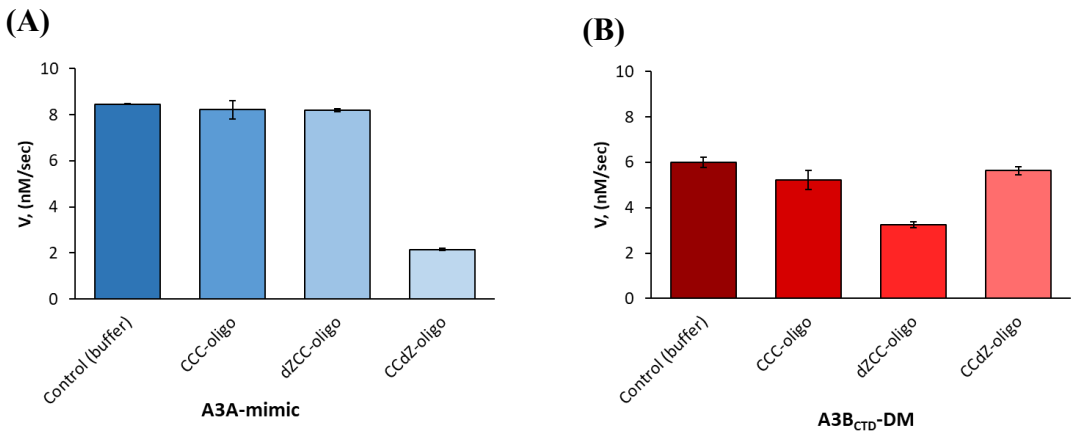


Figure 5.5 Screening of dZ-containing oligos for the selective inhibition of A3 enzymes.

Initial speed of the deamination (*V*) of 350 µM TCA-oligo (substrate) in the absence (buffer) and presence of 50 µM oligos (dZCC-oligo, CCdZ-oligo, or CCC-oligo control) by (A) 50 nM A3A-mimic (blue) in citrate-phosphate pH 5.5 buffer or (B) 2 µM $A3B_{CTD}$-DM (red) in citrate-phosphate pH 7.5 buffer measured at 298 K. Experiments were repeated twice and the mean value was plotted with error bars reported as standard deviations (SD).

To confirm that these selective inhibitors are not further processed by the inhibited enzymes, we recorded a ^{1}H-NMR spectrum of the CCdZ-oligo in the presence of A3A-mimic, which is the most active deaminase in our A3 variants. Examination of the NMR spectrum over ~ 4 hours revealed that no signals change, but importantly the other non-preferred dCs in the motif (CCdZ) were not deaminated as displayed by the lack of the dU signals in the range of 5.5 - 6.0 ppm (Figure 5.6). These observations confirm that the A3A-mimic strongly binds dZ in the CCdZ-oligo and is inhibited. Taken together, these results provide proof of concept that selective inhibition of A3 catalytic activity is feasible by using dZ instead of dC in different positions in a suboptimal DNA substrate containing the CCC motif.

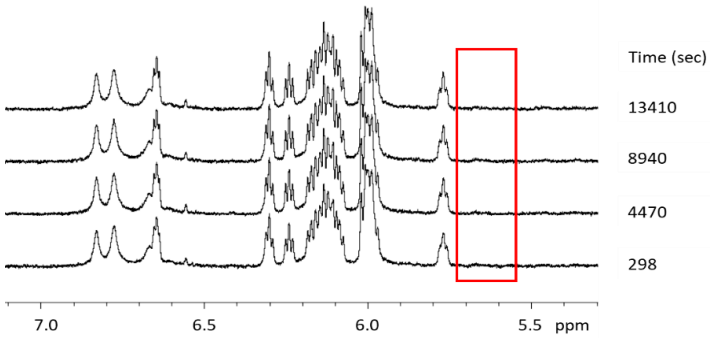


Figure 5.6 ^{1}H-NMR spectrum of CCdZ-oligo in the presence of A3A-mimic at different time points.

Experiment conducted using 350 µM CCdZ-oligo with 50 nM A3A-mimic in citrate-phosphate pH 5.5 buffer. No change is detected in CCdZ-oligo. Uracil product peaks should appear in the boxed region.

5.3.2 Quantitative characterisation of selective inhibition of A3A-mimic and A3B$_{CTD}$-DM using ^{1}H-NMR inhibitor assay

Next, we wanted to quantitatively characterise the extent of the inhibition of our A3A-mimic and A3B$_{CTD}$-DM by the aforementioned selective dZ-inhibitors (CCdZ-oligo and dZCC-oligo, respectively). The deamination of the TCA-oligo by our A3 variants was monitored using the previously described ^{1}H-NMR-based-inhibiton assay (see section 2.5.3) [94] in the presence of various concentrations of CCdZ-oligo or dZCC-oligo (2 – 100 µM). The inhibition constants (K_i, Figure 5.7 and Table 5.4) were derived from analysis of the linear dependence of the inverse deamination speed on inhibitor concentration based on a competitive mode of inhibition [135] (as described in section 4.4.2). As mentioned earlier, fitting the data using non-linear least squares, we obtain similar parameter values consistent with linear analysis (see appendix section Figure 8.7 and Table 8.4). We found that the K_i values of the selective dZ-oligos on our A3 variants (K_i

~12 µM for A3A-mimic CCdZ-oligo, and K_i ~ 19 µM for A3B$_{CTD}$-DM dZCC-oligo, Figure 5.7 and Table 5.4) were approximately two-fold weaker than the K_i values of our previously reported non-selective dZ-oligo inhibitor [135] (see Table 5.4). This is consistent with the two-fold weaker K_m's of the CCC-oligo of these enzymes in contrast to the K_m's of the preferred TCA-oligo (reported in Table 5.2). Nonetheless, we still observe a thirty-fold increase in the apparent affinity (K_m/K_i) when dZ is placed in the ssDNA sequence instead of the preferred dC for A3B$_{CTD}$-DM and A3A-mimic, as reported earlier in section 4.4.2 for the preferred TCA-motif. Although the CCdZ-oligo inhibited both the A3A-mimic and A3G$_{CTD}$, we noted that the A3A-mimic had a 2.8 times higher affinity (K_i ~ 19 µM) for this oligonucleotide in comparison to the A3G$_{CTD}$ (Ki ~ 53 µM) indicating the potential for further inhibition selectivity towards the A3A-mimic over the A3G$_{CTD}$ (Table 5.4).

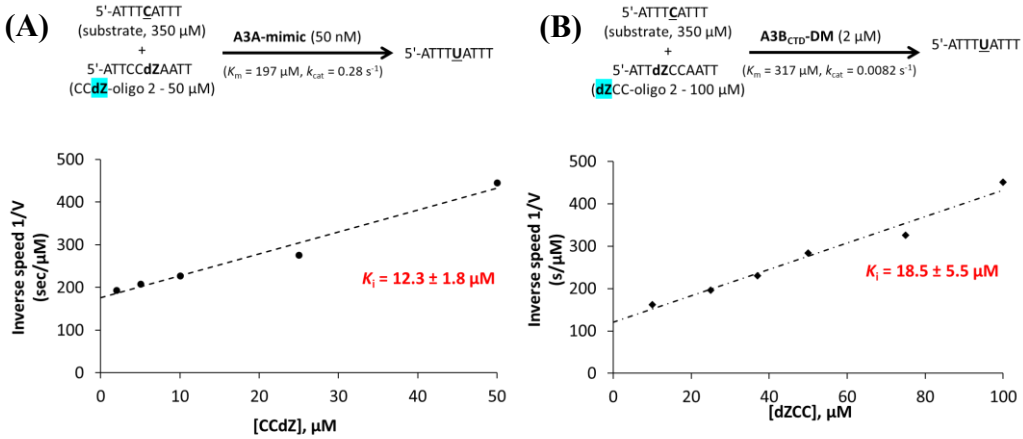


Figure 5.7 Quantitative characterisation of our A3 variants inhibition by the CCdZ-oligo and dZCC-oligo.

Inverse deamination velocity of 350 µM TCA-oligo by A3 variants in the presence of varying concentrations of selective dZ-inhibitors measured using ^{1}H-NMR at 298 K. (A) 50 nM A3A-mimic with CCdZ-oligo (2-50 µM) in citrate-phosphate pH 5.5 buffer, or (B) 2 µM A3B$_{CTD}$-DM with dZCC-oligo (2-100 µM) in citrate-phosphate pH 7.5 buffer. Experiments were fitted with linear regression to derive K_i.

Table 5.4 Inhibition constants (K_i, µM) for our A3 variants derived from ^{1}H-NMR-based inhibition assay.

Oligonucleotides	A3A-mimic K_i, µM	$A3B_{CTD}$-DM K_i, µM	$A3G_{CTD}$ (conducted by Dr. Stefan Harjes [135]) K_i, µM
dZ-oligo	7.5 ± 1.7 [135]	11.4 ± 2.6 [135]	Not an inhibitor *
CC**dZ**-oligo	12.3 ± 1.8	Not an inhibitor *	53 ± 10 [135]
dZCC-oligo	Not an inhibitor *	18.5 ± 5.5	Not an inhibitor

* Not an inhibitor means that dZ is not incorporated in the preferred substrate sequence or dC deamination target.

Lastly, we wanted to assess the deamination of the A3 enzymes preferred dC in these inhibitors to test the stability of the inhibitors against other A3 enzymes. Using the dZC<u>C</u>-oligo and <u>C</u>CdZ-oligo as substrates we determined the kinetic parameters of the A3A-mimic, $A3B_{CTD}$-DM (refer to Figure 8.3), and $A3G_{CTD}$ (experiments on $A3G_{CTD}$ conducted by Dr. Stefan Harjes [8]) as described earlier in section 5.2.2 using the ^{1}H-NMR based-activity assay (section 2.5.2).

Table 5.5 Kinetic parameters for A3 enzymes deaminating dC in different DNA substrates .

Oligonucleotides	A3A-mimic K_m, µM (k_{cat}, s^{-1})	$A3B_{CTD}$-DM K_m, µM (k_{cat}, s^{-1})	$A3G_{CTD}$ (conducted by Dr. Stefan Harjes [8, 94]) K_m, µM (k_{cat}, s^{-1})
T**<u>C</u>**A-oligo [135]	197 ± 30 (0.28 ± 0.041)	317 ± 55 (0.0082 ± 0.0014)	Not a substrate *
CC**<u>C</u>**-oligo	440 ± 80 (0.38 ± 0.06)	Not a preferred substrate +	570 ± 90 (0.10 ± 0.04) [94]
<u>C</u>CC-oligo	Not a preferred substrate +	590 ± 90 (0.009 ± 0.001)	Not a preferred substrate +
dZC**<u>C</u>**-oligo	520 ± 90 (0.44 ± 0.06)	Not a substrate *	≈ 1 mM (≈ 0.1) [8]
<u>C</u>CdZ-oligo	Not a substrate *	1370 ± 270 (0.010 ± 0.002)	Not a substrate *

***** Not a substrate means that none of dCs is deaminated under the experimental conditions reported.

+ Not a preferred substrate means that other dCs are also deaminated but at later time points when the concentration of the substrate with the preferred dC in the sequence decreased significantly due to deamination.

The dZCC-oligo, which selectively inhibits A3B$_{CTD}$-DM, was found to be a poor substrate for the A3A-mimic and A3G$_{CTD}$ [8] in comparison to the TCA-oligo and CCC-oligo (Table 5.5). More evidently, the CCdZ-oligo, which inhibits A3A-mimic and A3G$_{CTD}$ [135], was discovered to be a very poor substrate for A3B$_{CTD}$-DM with a K_m ~ 1370 μM (Table 5.5). Upon inclusion of dZ in the CCC motif, we observed that the A3 enzymes still deaminated their preferred dC (observed in Figure 8.3), but had much weaker apparent binding affinities K_m (Table 5.5). However, when we compared the K_m value of the dZCC-oligo deaminated by A3A-mimic (K_m ~ 520 μM, as it had the lowest K_m value of the dZ-oligos as a substrate, Table 5.5) to the K_i value for A3B$_{CTD}$-DM of the same oligo (dZCC K_i ~ 19 μM, Table 5.5), we found that this oligonucleotide had a 30 fold higher apparent affinity (K_m/K_i ~ 30) towards the A3B$_{CTD}$-DM than to other enzymes. Taken together, this would signify that inhibition of the A3B$_{CTD}$-DM by the dZCC-oligo would predominate over the deamination of the preferred dC in this oligonucleotide (dZCC-oligo) by the A3A-mimic, as well as the A3G$_{CTD}$. This conclusion will be valid for all other dZ-containing oligos (Table 5.4 and Table 5.5).

5.4 Conclusions

Taking advantage of the reported selective deamination of the wild-type A3B$_{CTD}$ and wild-type A3A on a dC in a suboptimal substrate containing several cytosines (CCC-motif) [37], we predicted that selective inhibition of the cancer-associated A3B would be achievable. We hypothesised that incorporation of our dZ nucleoside inhibitor in place of the preferentially deaminated dC in the CCC-motif would selectively inhibit these A3 enzymes. To test our hypothesis, we used our active A3B$_{CTD}$ variants, which showed that our A3B$_{CTD}$-DM preferred to deaminate the dC at the 5'-end of the 5'-CCC-oligo (K_m ~ 600 μM) like the wild-type A3B$_{CTD}$ [37], whereas our A3A-mimic (A3B$_{CTD}$-AL1) deaminated the dC at the 3'-end of the 5'-CCC-oligo (K_m ~ 400 μM) like the wild-type A3A [37] and A3G$_{CTD}$ [47-51].

These observations are consistent with our hypothesis, we found that placement of dZ near the 3'-end of the 5'-CCC-oligo selectively inhibited our A3B$_{CTD}$-DM variant (dZCC-oligo K_i ~ 19 μM), but not our A3A-mimic or the reported A3G$_{CTD}$ [135]. On the other hand, placement of dZ at the 5'-end of the motif specifically inhibited the A3A-mimic (CCdZ-oligo K_i ~ 12 μM) and a reported A3G$_{CTD}$ (CCdZ-oligo K_i ~ 53 μM) [135], but not A3B$_{CTD}$-DM. Notably, the A3A-mimic and A3G$_{CTD}$ could still deaminate their preferred 3'-dC in

the A3B$_{CTD}$-DM-selective dZCC-oligo inhibitor of (dZCC-oligo, A3A-mimic K_m ~ 520 μM and A3G$_{CTD}$ K_m > 1 mM), and the A3B$_{CTD}$-DM could also deaminate the preferred 5'-dC in the CCdZ-oligo inhibitor of A3A-mimic and A3G$_{CTD,}$ but with very weak K_m (CCdZ-oligo, A3B$_{CTD}$-DM K_m ~ 1.37 mM). However, as indicated by the K_i values mentioned earlier, these enzymes would bind to these selective inhibitors with a strong affinity relative to K_m, supporting a scheme where the A3 enzymes would bind to dZ rather than to any dC in the oligonucleotide and selective inhibition would overrule. Embedding dZ in place of preferentially deaminated dC in the CCC-oligo led to the first selective inhibitor of the A3B$_{CTD}$. The use of more potent inhibitors of cytosine deamination would enhance the inhibition effect for use in the cellular studies.

6. Evaluation of A3-ssDNA complexes using small-angle X-ray scattering

6.1 Introduction

Some of the work reported in this chapter was presented in several conferences listed below.

1) Barzak, F. M., Mohammadzadeh N., Kvach, M. V, Harjes S., Kurup, H. M., Aihara, H., Harris R. S., Krause, K. L., Chelico, L., Filichev, V. V., Harjes, E., Jameson G.B. Small-angle X-Ray scattering structures of human APOBEC3 bound to single-stranded DNA. Selected poster presentation at the Australian Synchrotron User Meeting 2018 held between the 22nd - 23rd November 2018 at the Australian Synchrotron, Melbourne, Victoria, Australia.

2) Barzak, F. M., Mohammadzadeh N., Kvach, M. V, Harjes S., Kurup, H. M., Aihara, H., Harris R. S., Krause, K. L., Chelico, L., Filichev, V. V., Harjes, E., Jameson G.B. Small-angle X-Ray scattering structures of human APOBEC3 bound to single-stranded DNA. Invited oral presentation at the Biodiscovery & Biochemistry Conference 2019 held between the 2nd - 4th July 2019 at the Victoria University, Wellington, New Zealand.

3) Barzak, F. M., Mohammadzadeh N., Kvach, M. V, Harjes S., Kurup, H. M., Aihara, H., Harris R. S., Krause, K. L., Chelico, L., Filichev, V. V., Harjes, E., Jameson G.B. Model structures of human APOBEC3 bound to ssDNA-based inhibitors using SAXS. Invited oral presentation at the 24th International Analytical Ultracentrifugation Workshop and Symposium held between the 24th - 29th August 2019 at the University of Canterbury, Christchurch, New Zealand.

Crystal structures have shed some light on single-domain inactive A3 bound to single-stranded nucleic acid substrates [129, 161, 239]; however, so far there has not been any experimentally based models and structures of catalytically active A3 bound to ssDNA or full-length two-domain enzymes with ssDNA. Several studies, including this study, reported that ssDNA substrates bind weakly to inactive A3 catalytic domains (medium-high μM range) [36, 94, 135, 172]. Therefore, in an attempt to enhance the binding affinity and prevent deamination complications, we used our recently characterised A3 inhibitors containing dZ-modified ssDNA, which had with low μM affinity, to study active A3 enzymes. Owing to challenges faced during crystallisation, small-angle X-ray scattering (SAXS) was used to model the active $A3B_{CTD}$-ssDNA complex and the two-domain A3-ssDNA complex.

SAXS operates by registering X-ray scattering pattern intensities of a sample in solution on a detector (Figure 6.1) [168]. The raw two-dimensional scattering data are processed through a number of steps. First, the raw scattering data are reduced (SCATTERBRAIN 2.82), then the data are plotted as a SAXS profile of integrated intensity versus frame number (CHROMIX, ATSAS 2.8.3 suite [75]). The frames of interest are averaged and buffer subtracted, transforming the data to a 1D-scattering curve where log of the scattering intensity is plotted against the scattering vector (log(*I*) vs *q*) (PRIMUSQT, ATSAS 2.8.3 suite [75]) (Figure 6.1) (described in materials and methods 2.6.2).

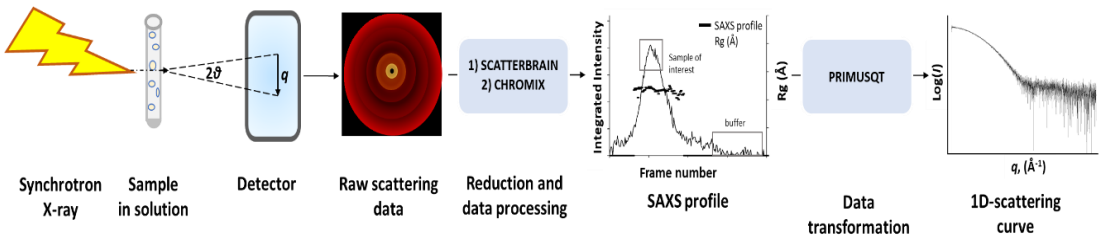


Figure 6.1 Schematic representation of a SAXS experiment and data processing.

The scattering vector *q* is defined as $q = 4\pi\sin\theta/\lambda$, 2θ is the scattering angle, and λ is the wavelength of 1.0332 Å for collected data.

Biophysical parameters such as the radius of gyration, maximum dimension, and molecular weight of the protein can be derived from the scattering curve to define the samples' morphology in solution.

6.1.1 Analysis of SAXS data

Standard SAXS experiments, termed batch mode SAXS, function by exposing the sample to monochromatic synchrotron X-rays and then collecting scattering data using an automated sample changer [168, 217]. This mode is beneficial to screen samples using low sample concentrations (between 0.25 mg/mL and ~10 mg/mL), run experiments using a range of temperatures (8-45 °C), use smaller sample volumes (40 µL), and retrieve samples following the experiment. Batch mode SAXS assumes that samples are monodispersed, and background scattering is accounted for by careful buffer matching. However, as there is potential of biological macromolecules to aggregate or form higher order structures in solution, interpretation of data derived using standard SAXS methods maybe hindered. Therefore, the use of SEC-FPLC directly coupled with SAXS (SEC-SAXS) resolves sample species and buffer matching, while additionally monitoring the elution profile by UV detection [218]. Furthermore, the use of a co-flow system along with SEC-SAXS (SEC-co-flow-SAXS) prevents sample degraded due to radiation damage

from coating the capillary walls and enhances the signal to noise ratio. In co-flow the sample flows in the centre of the cell while a matched buffer encases the sample and prevents protein denatured by the intense X-radiation depositing on the cell walls [117].

During SEC-co-flow-SAXS, the sample's elution profile is monitored by uv-visible spectroscopy at various wavelengths over time, while near-simultaneously collecting SAXS data of the sample per frame. However, a slight lag of 1.05 seconds/frame occurs between monitoring the elution profile followed by SAXS of the sample; this lag is accounted for during analysis. If examining the elution profile shows that the sample elutes as a homogeneous species off the SEC column, then the scattering data can be transformed into a 1D-scattering curve as described earlier in Figure 6.1 and analysis can proceed straightforwardly (Figure 6.2). However, if the sample is partially unresolved multiple species, then its scattering data must be first deconvoluted before proceeding with further analysis. Simple deconvolution involves using SVD/EFA BioXTAS RAW [105], where the singular value decomposition (SVD) function defines the number of components in the sample (referred to as eigenvalues). Then the evolving factor analysis (EFA) method can be used to define the boundaries and extract the scattering curves of each component (further described in Figure 8.9). However, for complicated poorly resolved samples Gaussian decomposition allows the deconvolution of the data and the extraction of the distinct scattering curves of each component (US-SOMO HPLC-SAXS module) [31-33].

Analysis was conducted using programs within the PRIMUSQT ATSAS 2.8.3 suite [75]. Initially, the 1D-scattering curve is converted into a double logarithmic plot to confirm that the sample assessed is homogeneous with a single species, confirmed by the presence of a plateau at lower scattering angles (q) values (Figure 6.2E). Plotting low q data using the Guinier distribution analysis ($\log(I)$ vs q^2) through the AUTORG method (Figure 6.2A) allows evaluation of the monodispersity of the sample in solution. This also provides an initial estimate of the radius of gyration (Rg) and the extrapolated intensity at zero scattering angle $I(0)$ describes the overall size of the molecule. The assumption of globular shape for the Guinier plot is valid when $qR\text{g} \leq 1.3$ (denoted $q.R$g max) Next, the data are converted into a Kratky plot (I^*q^2 vs q) to assess the shape and fold of the molecule (Figure 6.2B). In addition, the Kratky plot provides information regarding the oligomeric state of the molecule.

An indirect inverse Fourier transformation of the scattering data performed using AUTOGNOM results in the pairwise distribution function $P(r)$ curve, which represents the distribution of interatomic distances (r) within the molecule. The molecule's maximum diameter (D_{max}) can be determined from the $P(r)$ curve as $P(r)$ approaches zero at $r >> 0$ (Figure 6.2C). Furthermore, the Rg and $I(0)$ can be accurately calculated from the $P(r)$ curve using all the experimental data, unlike the Guinier analysis, which uses only a small subset at low q. The excluded particle volume (also termed Porod volume, V) is calculated through the DATPOROD program using $I(0)$ values attained from the $P(r)$ plot (Figure 6.2D). The Porod volume can then be used to directly compute the molecular weight (MW in Daltons) of the solute ($MW \approx V*0.6$) providing valuable information about the oligomeric state of the molecule (Figure 6.2). Furthermore, the MW can also be calculated from $I(0)$ if the concentration is accurately known using a previously described method [178].

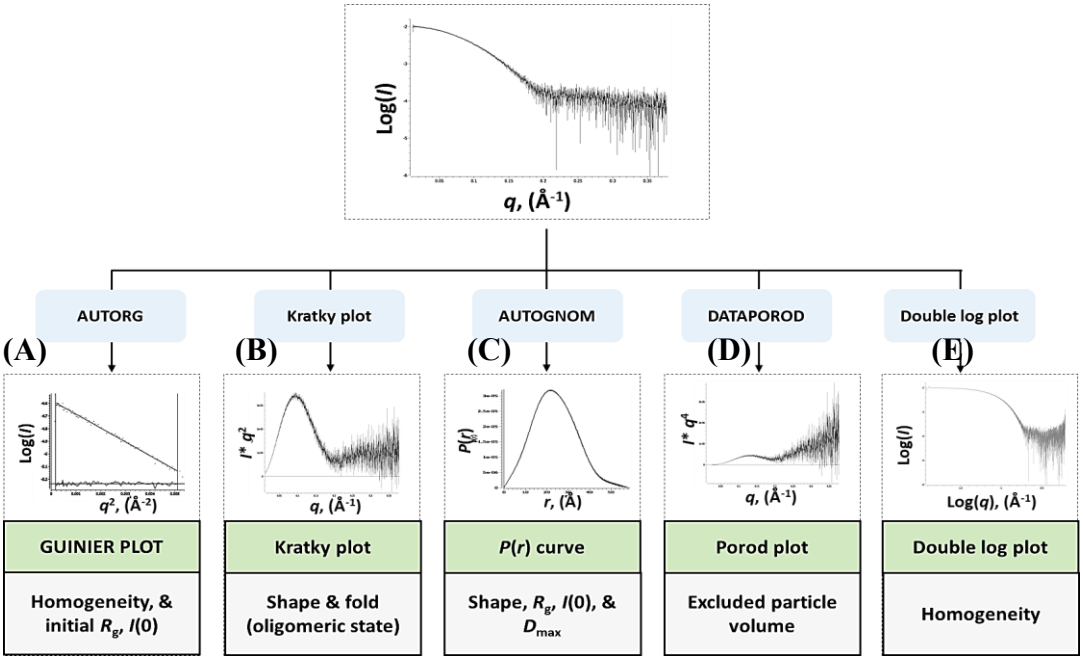


Figure 6.2 Schematic of the analysis of SAXS scattering data.

Low-resolution 3D-models can then be computed through *ab initio* shape restoration using the DAMMIF program by applying restraints of biophysical parameters attained from the 1D scattering curve. The DAMMIF program first generates several models, which are then averaged (DAMAVER) and are further filtered by cut-off volume constraints based on derived SAXS analysis parameters (DAMFILT) to produce the dummy filled models. The normalised spatial discrepancy (NSD) score quantitatively measures the similarities between the generated set of 3D envelope models, where NSD

≤ 0.8 is an acceptable variance. These envelope models can be superimposed with atomic models to assemble a high-resolution model. Lastly, rigid body modelling is conducted using FoXS [74, 226] and CRYSOL [75] by comparing the experimental scattering data to the back-calculated 1D-scattering profiles of atomic structures or model structures, to validate model. From these programs a fitting parameter termed the Chi^2 value (also called χ^2) can be obtained. Chi^2 gives a measure of discrepancies between the experimental scattering data to the back-calculated 1D-scattering profiles of atomic structures; a Chi^2 equal to one would indicate a perfect fit.

6.2 Active $A3B_{CTD}$-inhibitor ssDNA SAXS model structure

Until 2016, the structure of the $A3B_{CTD}$-ssDNA complex was unknown. Structural studies were conducted using an inactive $A3B_{CTD}$-QM-ΔL3-A3Aloop1swap-E255A chimera ($A3B_{CTD}$*) along with a 5mer ssDNA substrate oligonucleotide yielding the first visualisation of the $A3B_{CTD}$*-ssDNA complex (PDB: 5TD5) [239]. Unlike this study [239], our study focused on using SAXS to examine an active $A3B_{CTD}$ bound to our ssDNA-based inhibitor to visualise the relative complex structure in solution. Using this method we can compare our SAXS model structure to the published inactive $A3B_{CTD}$-ssDNA crystal structure (PDB: 5TD5) [239]. SAXS experiments were performed using our previously described active $A3B_{CTD}$-AL1 variant ($A3B_{CTD}$-QM-ΔL3-AL1swap) along with our previously characterised dZ-containing oligo (see chapter 4) to prevent cytosine deamination in the substrate, enhance binding to the protein, and to examine the formation of this complex. This oligonucleotide had 9 base pairs with dZ (see section 2.3.2) as at the time 5FdZ had not yet been characterised (Figure 4.9).

6.2.1 Sample preparation of $A3B_{CTD}$

Initial screenings were conducted using batch mode SAXS with varying concentrations of $A3B_{CTD}$-AL1 (1, 2.5, 5, 10 mg/mL) to gain insight of the relative biophysical characteristics, especially of reversibly associating dimeric or higher oligomeric species. To ensure a homogeneous sample, and reduce background signal noise, SEC-co-flow-SAXS was exploited to examine the samples in solution. Preliminary analysis of the $A3B_{CTD}$-DM protein resulted in a poor scattering curve as a consequence of a large proportion of protein aggregating on the SEC column. Conversely, the $A3B_{CTD}$-AL1 eluted from the column as a single peak producing a good scattering curve (Figure 6.3A).

Therefore, this variant was selected to study its formation in solution with and without dZ-containing ssDNA.

To assess the protein's stability, 10 mg/mL of A3B$_{CTD}$-AL1 was examined using SEC-FPLC in the previously mentioned 50 mM citrate-phosphate pH 5.5 buffer (termed kinetic pH 5.5 buffer in this chapter) and the high salt pH 7.4 buffer (see Table 2.15). As demonstrated earlier, the A3B$_{CTD}$-AL1 variant deaminates target dC within ssDNA more efficiently in the mildly acidic pH 5.5 buffer (see Figure 4.7), as this pH influences interactions between the protein and ssDNA. However, to maintain the protein's stability, A3B$_{CTD}$ proteins prefer the high salt pH 7.4 buffer (refer to section 3.2.3). Therefore, buffer choice was an important consideration to allow both the formation of a strong A3B$_{CTD}$-ssDNA complex and also to prevent protein aggregation during SAXS experiments.

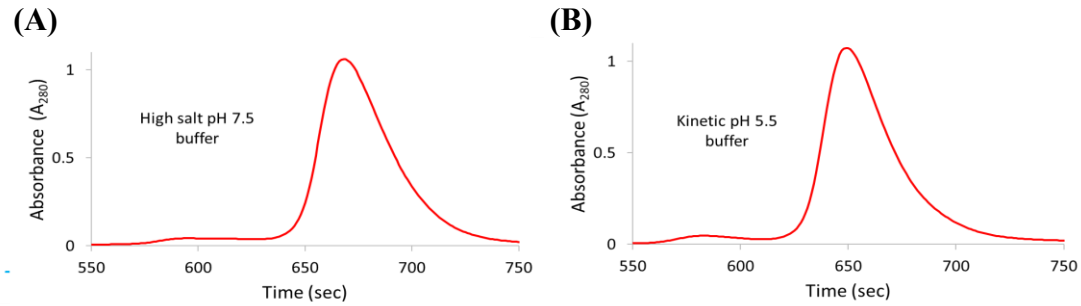


Figure 6.3 Size exclusion chromatography elution profiles of A3B$_{CTD}$-AL1 in varying buffers. SEC-FPLC performed using (A) A3B$_{CTD}$-AL1 (10 mg/mL) in high salt pH 7.4 buffer (50 mM Tris-HCl, pH 7.4, 0.5 M NaCl, and 2 mM β-ME), (B) A3B$_{CTD}$-AL1 (10 mg/mL) in kinetic pH 5.5 buffer (50 mM citrate-phosphate, pH 5.5, 0.2 M NaCl, 2 mM β-ME). Samples were run at 25 °C through a SEC column at a flow rate of 0.2 mL/min and monitored using wavelengths of 280 nm (A_{280}).

We established that our purified monomeric A3B$_{CTD}$-AL1 elutes predominantly as a single peak in the same manner in both kinetic and high salt buffers as observed in Figure 6.3. A3B$_{CTD}$-AL1 peak maximum at around 668 seconds in the high salt pH 7.4 buffer (Figure 6.3A), whereas in kinetic pH 5.5 buffer the protein elutes slightly earlier with a peak maximum at 650 seconds (Figure 6.3B). These minor differences in retention times are attributed to the variations in buffer composition, which may affect the properties of the proteins such as the charge of the residues and hydrophobicity, therein influencing the separation by SEC-FPLC. However, as there were no obvious signs indicating protein aggregation occurring in the kinetic pH 5.5 buffer during SEC-FPLC elution, this buffer

was selected for our $A3B_{CTD}$ SAXS experiments. Using the kinetic pH 5.5 buffer will drive the formation of the $A3B_{CTD}$-ssDNA complex as described earlier in Figure 4.7.

6.2.2 Evaluation of the $A3B_{CTD}$-AL1 and $A3B_{CTD}$-AL1/ssDNA complex using SEC-SAXS

Scattering data of the $A3B_{CTD}$-AL1 (5 mg/mL), dZ-oligo, and $A3B_{CTD}$-AL1 along with dZ-oligo at varying ratios were acquired in the kinetic pH 5.5 buffer (containing 10% glycerol to prevent radiation damage). In conjunction, UV elution profiles were collected to aid with the visualisation of the eluent position off the SEC column. Initial assessment of the elution profiles showed that the $A3B_{CTD}$-AL1 eluted as a single peak with a peak maximum at around 648 seconds (Figure 6.4A), whereas the dZ-oligo peak maximum was at around 662 seconds (Figure 6.4B). Overlaying the trace *R*g values over the elution profile revealed that the protein had an average *R*g ~ 18.7 Å, while the dZ-oligo had an *R*g ~ 9.9 Å. Moreover, both samples were monodisperse as indicated by the constant plateau of *R*g values across the elution peak seen in Figure 6.4A and B.

Initial screenings were performed to determine the ratio of protein to oligonucleotide required to ensure the $A3B_{CTD}$-AL1-ssDNA complex had formed, as monitored by UV. Comparing the elution peak maximum of the $A3B_{CTD}$-AL1 (blue arrow, Figure 6.4A) to the peak maximum of the $A3B_{CTD}$-AL1 in the presence dZ-oligo at a ratio of one to one (protein to ssDNA) (highlighted with blue bar, Figure 6.4C), we found an increase of ~ 0.3 absorbance units (Au) in the protein peak maximum (blue arrow, Figure 6.4C). This increase directly correlated with a decrease of ~ 0.3 Au of the dZ-oligo peak maxima (green arrow in Figure 6.4B-C). The change in the absorbance implied that the oligo was bound to the $A3B_{CTD}$-AL1. These observations are consistent, upon binding of the oligo to the $A3B_{CTD}$-AL1, the *R*g of the protein increased in size from *R*g ~ 18.7 Å (Figure 6.4A) to *R*g ~ 19.5 Å (Figure 6.4C-E), which also caused the protein to elute off the column earlier (after ~ 620 seconds, Figure 6.4C-E), further supporting the formation of the $A3B_{CTD}$-AL1/dZ-oligo complex.

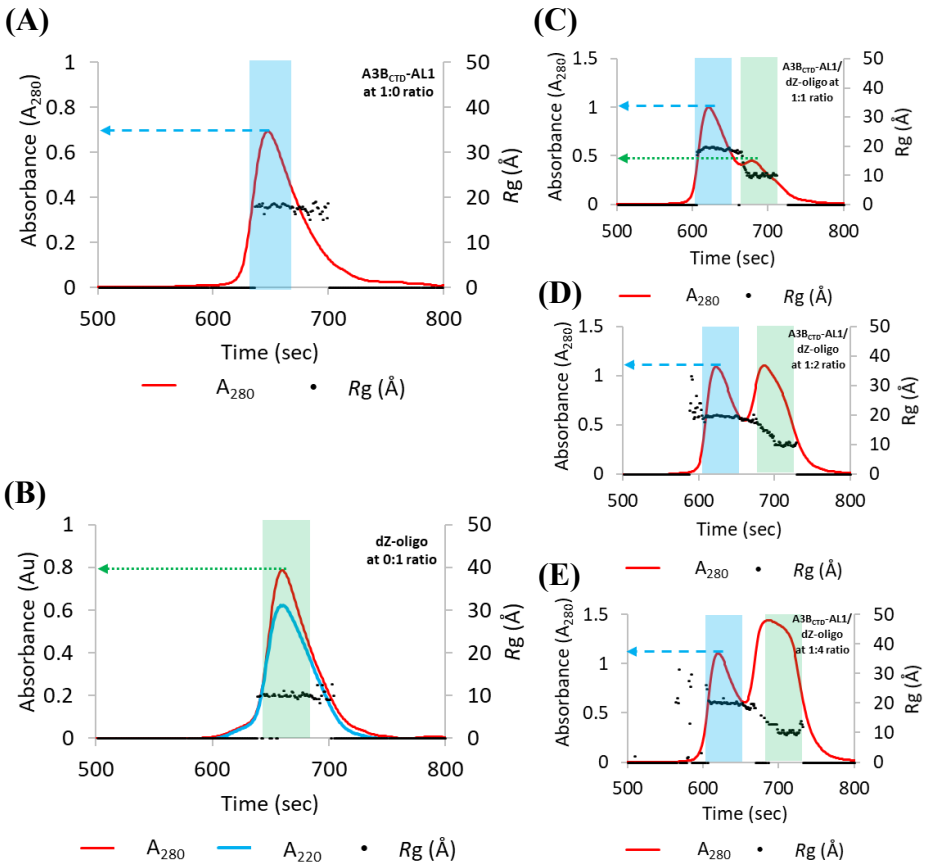


Figure 6.4 SEC elution profiles of A3B$_{CTD}$-AL1 in varying molar ratios with dZ-oligonucleotide. SEC elution profiles of (A) A3B$_{CTD}$-AL1 at 5 mg/mL (~ 227 µM), (B) dZ-oligo at 0.68 mg/mL (~ 226 µM), (C) A3B$_{CTD}$-AL1 with dZ-oligo at a 1 to 1 ratio, (D) A3B$_{CTD}$-AL1 with dZ-oligo at a 1 to 2 ratio, (E) A3B$_{CTD}$-AL1 with dZ-oligo at a 1 to 4 ratio. Samples were run at 25 °C through a SEC column at a flow rate of 0.2 mL/min in kinetic pH 5.5 buffer (50 mM citrate-phosphate pH 5.5, 0.2 M NaCl, 2 mM β-ME). Blue bar region and arrow represent A3B$_{CTD}$-AL1 elution, green bar region and arrow represents dZ-oligo elution. Ratio represents molar ratio of protein mixed with ssDNA.

Normally, to check for the presence of DNA along with protein elution we can take the maximum A$_{260}$ value and divide by the maximum A$_{280}$ value, but due to technical difficulty A$_{260}$ values were not collected for these samples. As mentioned earlier, in the presence of dZ-oligo, the protein peak maximum at a ratio of one to one (protein to oligo) increases by ~ 0.3 Au (blue arrow, Figure 6.4C), at ratio of one to two it is increased by ~ 0.4 Au (blue arrow, Figure 6.4D). Whereas, no further increase in peak maximum occurs at the one to four ratio (blue arrow, Figure 6.4E), indicating that the A3B$_{CTD}$-AL1 protein was saturated with the oligo at a ratio of one to two (protein to oligo) or higher. So, for further SAXS experiments using A3B$_{CTD}$-AL1, a one to two ratio of protein to oligo was selected to ensure formation of a complex with dZ-containing ssDNA.

6.2.3 Model SAXS structures of the active A3B$_{CTD}$-AL1 and A3B$_{CTD}$-AL1/dZ-oligo complex

Following SEC-co-flow-SAXS of A3B$_{CTD}$-AL1, dZ-oligo, and A3B$_{CTD}$-AL1/dZ-oligo (1:2 ratio) samples, the raw scattering data were processed and plotted as SAXS profile plots (integrated intensity versus the frame number) along with an *R*g trace (see Figure 6.5A). The dZ-oligo resulted in an *R*g ~ 9.5 Å (Figure 6.4B), consistent with having 9 nucleotides. However, due to its small size and limited resolution of SAXS (*R*g > 10 Å) further analysis was restricted and not continued for this sample.

On the other hand, both the A3B$_{CTD}$-AL1 and A3B$_{CTD}$-AL1/dZ-oligo SAXS data were further investigated. Based, on the steady *R*g across the elution profiles (blue bar region, Figure 6.4A and D) and scattering profiles of the A3B$_{CTD}$-AL1 (frames 610-660, Figure 6.5A) and A3B$_{CTD}$-AL1/dZ-oligo (frames 570-635, Figure 6.5A) we can conclude that a single homogeneous species is present. Therefore, these frames were averaged and transformed into 1D-scattering curves as described in section 2.6.2. From the 1D-scattering curves several biophysical structural parameters were extracted that describe the molecules' configuration in solution (listed in Table 6.1 and displayed in Figure 6.5).

Assessment of A3B$_{CTD}$-AL1 and the A3B$_{CTD}$-AL1/dZ-oligo complex SAXS data using the Guinier distribution showed that both samples had a good fit to the linear regression at low scattering angles, indicating that the samples were not aggregated and were monodispersed (Figure 6.5B i). Moreover, at low *q* values the double logarithmic plot plateaued confirming that the samples were homogeneous (Figure 6.5B ii). Estimation of *R*g and *I*(0) values from the Guinier slope agreed well with those calculated by the independent *P*(*r*) method for both samples (listed in Table 6.1), further confirming the quality and relative size of the solutes. The Kratky plot exhibited a bell-shaped peak at low *q* (peak maxima at *q* ~ 0.1 Å) that then converged to the scattering axis at the higher *q* range (Figure 6.5C), indicating that the molecules were well-folded and globular, with not much flexibility. As expected, the *R*g, D_{max}, and Porod volume increased slightly upon formation of the A3B$_{CTD}$-AL1/dZ-oligo complex in comparison to the ligand-free A3B$_{CTD}$-AL1 (as displayed in Table 6.1). Furthermore, the *P*(*r*) plot displayed a symmetrical histogram curve, which implied that the molecule (both with and without

dZ-containing ssDNA) forms a compact near-spherical shape, consistent with results observed using the Kratky plot (Figure 6.5C).

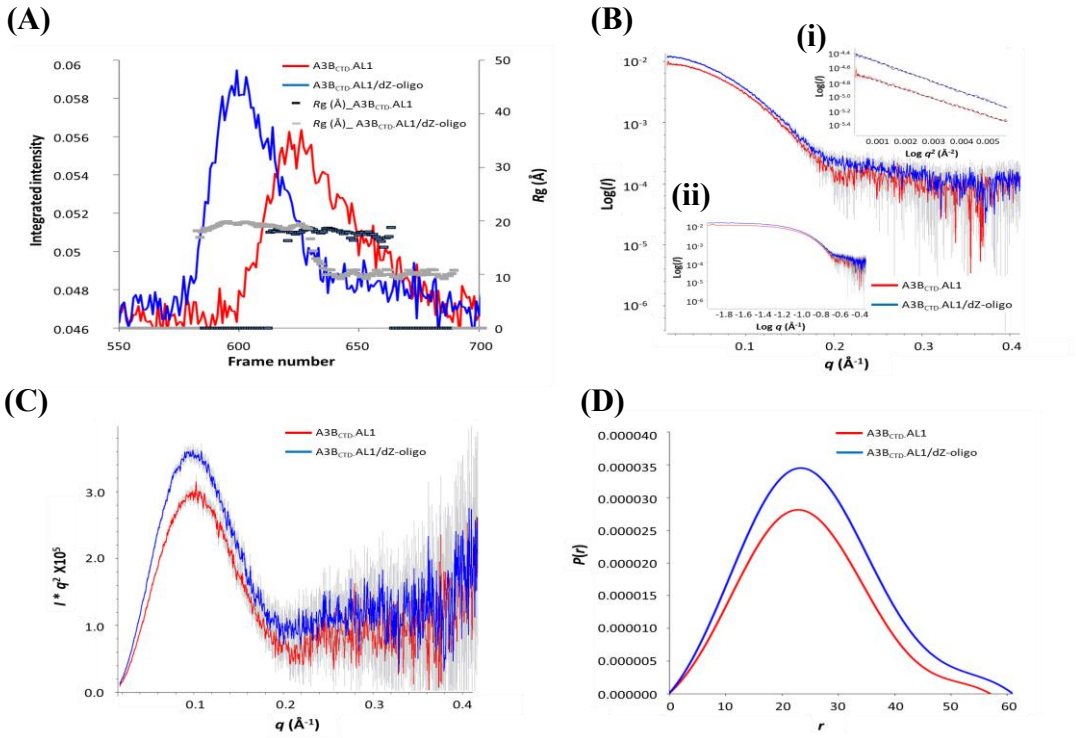


Figure 6.5 SEC-SAX analysis of A3B$_{CTD}$-AL1 and A3B$_{CTD}$-AL1/dZ-oligo complex.
SAXS scattering data of A3B$_{CTD}$-AL1 (red) and A3B$_{CTD}$-AL1/dZ-oligo (blue) samples were analysed using several programs from PRIMUSQT ATSAS 2.8.3 suite [75]. (A) SAXS profile plots, (B) scattering curves with (i) Guinier plot insert, (ii) double log plot insert, (C) Kratky plot, (D) *P*(*r*) distribution plot. Experiments conducted injecting into SEC-FPLC 5 mg/mL A3B$_{CTD}$-AL1 and A3B$_{CTD}$-AL1/dZ-oligo at a 1 to 2 molar ratio in kinetic pH 5.5 buffer.

These parameters were then used to calculate the molecular weight (MW). From the Porod volume a MW ~ 21 kDa and from *I*(0) a MW ~ 23 kDa were calculated for the A3B$_{CTD}$-AL1. The MW estimates were comparable to the known MW of A3B$_{CTD}$ (MW ~ 22 kDa) [239-240]. The A3B$_{CTD}$-AL1/dZ-oligo sample had slightly larger MWs and increase in the overall subunit ratio (summarised in Table 6.1), which further supports that a protein-ssDNA complex was formed.

Table 6.1 SAXS structural parameters of the A3B$_{CTD}$-AL1 and A3B$_{CTD}$-AL1/dZ-oligo complex

Structural parameters	A3B$_{CTD}$-AL1	A3B$_{CTD}$-AL1/dZ-oligo
Guinier analysis		
$I(0)$ (cm^{-1})	0.00940 ± 0.00005	0.01031 ± 0.00005
Rg (Å)	18.8 ± 0.2	19.8 ± 0.2
$q$min (Å^{-1})	0.0106	0.0106
$q \cdot R$g max	1.3	1.3
Coefficient of correlation (R^2)	0.98	0.99
P(r) analysis		
$I(0)$ (cm^{-1})	0.00930 ± 0.00004	0.01020 ± 0.00005
Rg (Å)	18.7 ± 0.1	19.7 ± 0.1
D_{max} (Å)	57	60
q range (Å^{-1})	0.0106 - 0.3539	0.0106 - 0.3539
Quality estimate	0.87	0.85
Porod volume (Å^3)	35,600	36,700
MW (Porod Volume*0.6) (Da) (ratio to expected subunit 22,000 Da)	21,360 (0.97)	22,020 (1.00)
MW (from $I(0)$ [178]) (Da) (ratio to expected subunit 22,000 Da)	23,330 (1.06)	25,310 (1.15)

Following parameter analysis, envelope models of the A3B$_{CTD}$-AL1 and the A3B$_{CTD}$-AL1/dZ-oligo-oligo complex were generated from the experimental scattering curves using the *ab initio* shape restoration programme (DAMMIF). Using parameter information derived from the scattering curves ten *ab initio* models were generated for each sample using *P*1 symmetry, as these samples were determined to be monomeric. The envelope models were filtered (DAMAVER and DAMFILT) into a single model to use for atomic modelling. A mean NSD score of 0.559 for the A3B$_{CTD}$-AL1 envelope and of 0.826 for the A3B$_{CTD}$-AL1/dZ-oligo complex envelope indicated a very good to acceptable consistency, respectively, for the generated ensembles (acceptable NSD ≤ 0.8). Table 6.2 summarises the envelope modelling.

Table 6.2 SAXS fitting and modelling parameters of the A3B$_{CTD}$-AL1 and A3B$_{CTD}$-AL1/dZ-oligo complex

Modelling parameters	A3B$_{CTD}$-AL1	A3B$_{CTD}$-AL1/dZ-oligo
Ab intio restoration		
DAMMIF (default parameters, 10 calculations)		
q range (Å^{-1}) for fitting	0.0106 - 0.3539	0.0106 - 0.3539
Symmetry	1 (none)	1 (none)
NSD (standard deviation)	0.559 (0.019)	0.862 (0.037)
Resolution (from SASRES) (Å)	28 ± 2	29 ± 2
Structure modelling		
PDB structure	5CQI	5TD5*
q range for modelling (Å^{-1})	0.0106 - 0.3020	0.0106 - 0.3020
FoXS, Chi^2	1.16	1.51
CRYSOL, Chi^2	0.81	0.86

Note: 5TD5* modified ssDNA length to 9mer.

The published X-ray crystal structures 5CQI [240] (A3B$_{CTD}$-QM-ΔL3-E255A) and 5TD5 [239] (inactive A3B$_{CTD}$-QM-ΔL3-A3ALoop1swap-E255A/5'-TTCAT complex) were utilised for manual docking of structures into the experimentally derived SAXS envelope models. However, to match the size of our dZ-oligo, four nucleotides were added onto the existing 5mer oligonucleotide in the 5TD5 structure (5'-ATTTCATTT) using PyMol [228] (termed 5TD5*).

The A3B$_{CTD}$-AL1 envelope model exhibited a globular shape with a slight concave groove (insert in Figure 6.6A). On the other hand, the A3B$_{CTD}$-AL1/dZ-oligo envelope had an additional unoccupied electron density that resembled an arm near and a less prominent groove. This empty electron density region was presumed to be part of the dZ-oligo indicating the location of the active-site. Superimposing our low-resolution dummy models with the atomic structures (5CQI or 5TD5* [239-240]) resulted in a good visual fit, indicating that the derived SAXS envelope models corresponded well to the published atomic structures [239-240]. Based on the superimposed DNA bound A3B$_{CTD}$ atomic structure (5TD5* [239]) and the unoccupied electron density observed in the envelope model of the A3B$_{CTD}$-AL1/dZ-oligo sample, we concluded that our dZ-oligo inhibitor bound to our A3B$_{CTD}$-AL1 protein in a similar manner as reported, where the ssDNA binds in a U-shaped conformation [239].

Interestingly, our dZ-oligonucleotide bound to the A3B$_{CTD}$-AL1 protein, but the 3'-end of the oligonucleotide seemed to be loosely attached and was fairly flexible, as indicated by the shape of its protrusion from the protein (Figure 6.6), consistent with reports of A3-ssDNA structures [93, 239]. Nevertheless, to validate our generated SAXS models, rigid body modelling was performed by comparing back-calculated 1D scattering profiles of the atomic structure models (PDB; 5CQI and 5TD5*) to the experimentally derived scattering data using FoXS [74, 226] and CRYSOL [75]. The scattering data of both samples showed a good fit to the X-ray models, as summarised in Table 6.2 and illustrated in Figure 6.6.

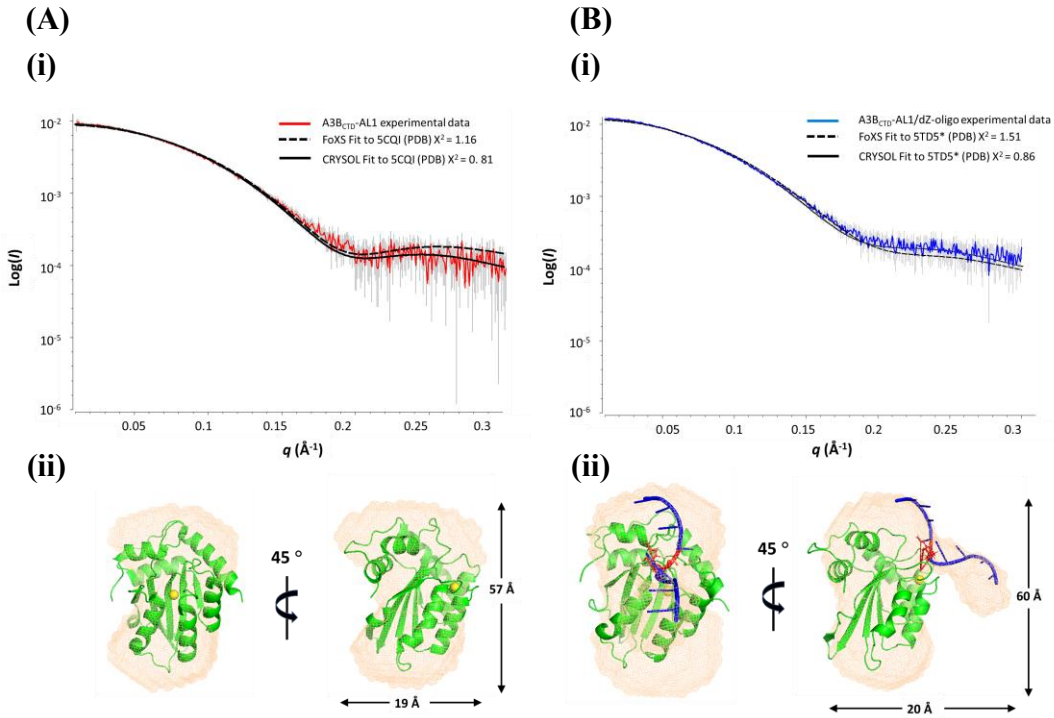


Figure 6.6 Fitting of A3B$_{CTD}$-AL1 SAXS profiles to known A3B$_{CTD}$ crystal structures.
(A) Model fit of A3B$_{CTD}$-AL1 to 5CQI crystal structure, (B) model fit of A3B$_{CTD}$-AL1/ dZ-oligo complex (injected into SEC-FPLC in a 1 to 2 ratio) to 5TD5* crystal structure.

In contrast, when we try to fit our ligand-free A3B$_{CTD}$-AL1 data to the ssDNA-bound A3B$_{CTD}$ atomic structure (5TD5*) or our A3B$_{CTD}$-AL1/dZ-oligo data to the ligand-free A3B$_{CTD}$ atomic structure (5CQI) we observe a poor fit (refer to Figure 8.8). Therefore, we can conclude that in solution based the SAXS models of A3B$_{CTD}$-AL1 and A3B$_{CTD}$-AL1/dZ- oligo complex are comparable to the X-ray crystal structures (PDB 5CQI and 5TD5, respectively).

6.3 SAXS analysis of the A3B$_{CTD}$-AL1 homodimer

6.3.1 Purification of the A3B$_{CTD}$-AL1 multimer species

During purification of the active A3B$_{CTD}$-AL1 protein, two predominant peaks eluted from the SEC column (Figure 6.7A). As mentioned previously in section 4.2.2, the peak eluting after 90 mL of buffer (termed peak 2) corresponded to the monomeric A3B$_{CTD}$-AL1 protein (used in kinetic studies in chapter 4 and SAXS modelling in section 6.2). To our surprise another peak (termed peak 1) eluted off the SEC column earlier, following 70 mL of buffer (Figure 6.7A). Fractions from both these peaks were visualised on a reducing SDS-PAGE gel revealing that they had the same molecular weight of ~22 kDa (Figure 6.7B), suggesting that both peaks contain the A3B$_{CTD}$-AL1 protein (non-reducing PAGE gel not performed due to time constraint). The protein eluting in peak 1 was presumed to be a multimer of A3B$_{CTD}$-AL1, as it is known that larger proteins elute first.

Mass spectroscopy analysis noted that the eluted protein from peak 1 corresponded to A3, while a previously described *in vitro* DNA deamination in-gel assay [240] verified that this protein was a catalytically active A3B$_{CTD}$ capable of cytosine deamination (refer to Figure 8.2). The SUMO-A3B$_{CTD}$-DM with a molecular weight of 38 kDa eluted off the same SEC column with a peak maximum located around 77 mL (refer to Figure 4.5). In comparison, the A3B$_{CTD}$-AL1 species from peak 1 eluted with a peak maximum at around 74 mL (Figure 6.5). Hence, we can propose that the A3B$_{CTD}$-AL1 protein from peak 1 would be almost of similar hydrodynamic size (around ~ 44 kDa) as the SUMO fused A3B$_{CTD}$-DM protein. We hypothesise that the A3B$_{CTD}$-AL1 from peak 1 may exist as a homodimer and this species will be further examined using SAXS (referred to as A3B$_{CTD}$-AL1 putative dimer in this section of this chapter).

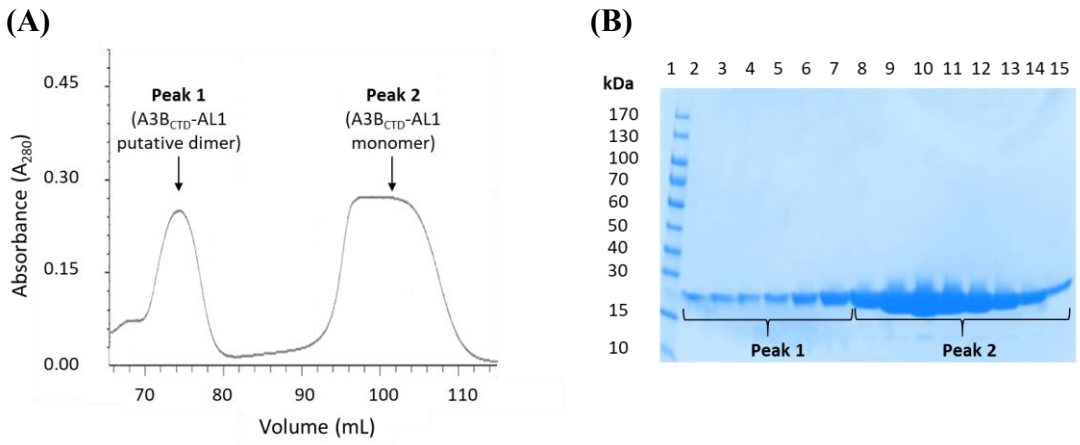


Figure 6.7 SEC-FPLC purification of the A3B$_{CTD}$-AL1 multimers.

SEC-FPLC of A3B$_{CTD}$-AL1 sample run through a SEC column at a flow rate of 0.5 mL/min in high salt pH 7.4 buffer (50 mM Tris-HCl, pH 7.4, 0.5 M NaCl, and 2 mM β-ME) at 25 °C. (A) SEC elution profile displaying the elution of the two forms of A3B$_{CTD}$-AL1, peak 1 and peak 2. (B) 16% Tricine SDS-PAGE gel (reducing conditions) of the purified A3B$_{CTD}$-AL1 fractions after SEC purification.

6.3.2 Evaluation of A3B$_{CTD}$-AL1 dimerisation

Several studies have described the prospect of multimerisation of the APOBEC proteins mediated through surface interactions [13, 23, 233, 276]. In this study, the A3B$_{CTD}$-AL1 protein was observed to elute as two species, a monomeric (peak 2) and a putative dimer (peak 1). In contrast, the other A3B$_{CTD}$ variants used during this research, A3B$_{CTD}$-QM-ΔL3, A3B$_{CTD}$-QM-ΔL3-E255A, and A3B$_{CTD}$-DM, all eluted as single peaks and were monomeric proteins (see sections 3.2.3 and 4.2.2). As A3B$_{CTD}$-AL1 adopts some characteristics of A3A and mimics its actions [37], it alludes to the notion that this observed dimerisation may be caused by the incorporation of the A3A loop 1 into A3B$_{CTD}$. Previous reports have suggested that A3A exists as both a monomer and a dimer *in vitro* [23, 129]; however, the significance of A3 dimerisation on the function and stability of the enzymes remains unclear. Therefore, examination of our A3B$_{CTD}$-AL1 putative dimer in solution may shed some light on its assembly and action.

6.3.2.1 SEC-FPLC assessment of the A3B$_{CTD}$-AL1 putative dimer

The presence of surface-exposed cysteine residues, Cys239 and Cys354, in A3B$_{CTD}$-AL1 raises the possibility of a disulfide-bridged homodimeric species forming under oxidising conditions (see section 6.3.1, where the reducing SDS-PAGE showed that the putative dimer travelled in a similar manner as the monomeric A3B$_{CTD}$-AL1 species). Initially, the A3B$_{CTD}$-AL1 putative dimer was refined through SEC-FPLC in reducing conditions. However, this larger species remained intact, signifying that under non-denaturing

conditions of SEC-FPLC (Figure 6.8A) (as contrasted with SDS-PAGE, Figure 6.7B) this conformation is stable and the disulfide bridge must be buried and inaccessible to the reducing agent.

To determine if the $A3B_{CTD}$-AL1 is in equilibrium between the two multimeric states (putative dimeric and monomeric forms), the dissociation behaviour of the $A3B_{CTD}$-AL1 putative dimer was examined. First, the $A3B_{CTD}$-AL1 putative dimer was concentrated to approximately 5 mg/mL, then examined using SEC-FPLC in both the standard high salt pH 7.4 buffer and kinetic pH 5.5 buffer. In both buffers the protein behaved in a similar manner, so for simplification results are presented for the high salt pH 7.4 buffer. The concentrated $A3B_{CTD}$-AL1 putative dimer sample eluted off the SEC column as two peaks. The majority of the sample ~ 93 % eluted off the column after 70 mL of buffer (peak 1) corresponding to the $A3B_{CTD}$-AL1 putative dimeric form, whereas a very minor peak of ~ 7 %, which correlated to the monomeric $A3B_{CTD}$-AL1 species, was also visible (peak 2, Figure 6.8A). In parallel, when the monomeric $A3B_{CTD}$-AL1 sample (10 mg/mL) was concentrated we observed a similar elution effect. Here, the majority eluted as a monomer (~ 96 %), while a very small percentage ~ 4 % eluted as the larger dimeric species (see small peak at 580 seconds, Figure 6.3). These findings suggested that an equilibrium between the putative dimeric and monomeric $A3B_{CTD}$-AL1 species, and vice versa, is unlikely.

To further explore this idea, the $A3B_{CTD}$ putative dimer was additionally studied under dilute concentrations (0.5, 1.25, and 2.5 mg/mL in high salt pH 7.4 buffer) to determine if the degree of dissociation of putative dimer changes as a function of concentration. Inspection of the elution profiles illustrated that for all the concentrations tested, the $A3B_{CTD}$-AL1 putative dimer resulted in the same relative amount of monomeric $A3B_{CTD}$-AL1 (7-8%, observed at peak maximum around 710 seconds, Figure 6.8B), comparable with previous observations shown in Figure 6.8A. As the assembly and disassembly of the $A3B_{CTD}$-AL1 putative dimer is not triggered by a concentration impact, and this putative dimer's stability is maintained in solution, it can be concluded that self-association of the $A3B_{CTD}$-AL1 monomeric subunits maybe facilitated by a disulfide linkage.

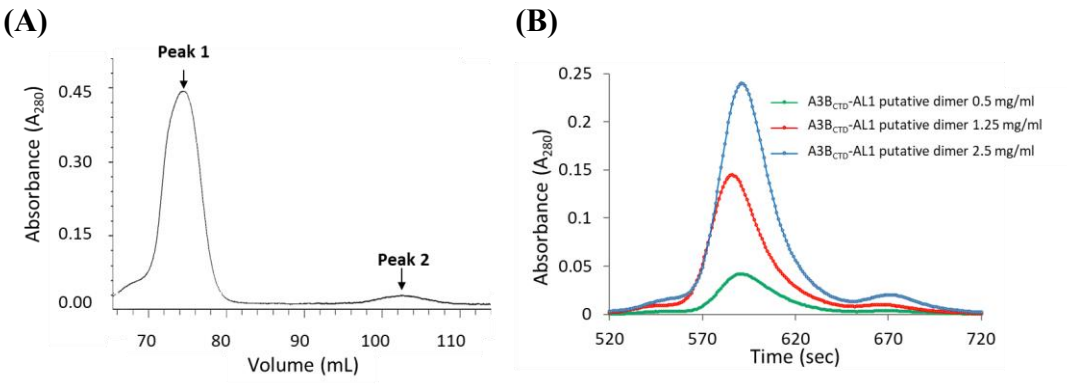


Figure 6.8 Purification of the A3B$_{CTD}$-AL1 putative dimer using SEC-FPLC.

Analysis of the A3B$_{CTD}$-AL1 putative dimer using SEC-FPLC. (A) SEC elution profile of 5 mg/mL A3B$_{CTD}$-AL1 putative dimer, (B) SEC elution profiles of the A3B$_{CTD}$-AL1 putative dimer at varying concentrations 0.5 - 2.5 mg/mL. SEC-FPLC was conducted at a flow rate of 0.5 mL/min through the SEC column in high salt pH 7.4 buffer (50 mM Tris-HCl, pH 7.4, 0.5 M NaCl, and 2 mM β-ME) at 25 °C.

6.3.2.2 Interface assessment of A3B$_{CTD}$-AL1 dimer

Assessment of the potential binding interface region between two A3B$_{CTD}$ subunits was investigated using the PRISM 2.0 webserver [9, 264]. This server uses a prediction-based algorithm with known interacting protein complexes as the template interface dataset. The target protein's surface is first structurally aligned to template interfaces, then hotspot binding regions on the target protein are identified. Using this information, the proteins are modelled onto the selected template using docking methods, then rigorous refinement is conducted to predict the interface complex of the target proteins. The favourability of a given binding reaction is described by Gibbs free energy (ΔG), where ΔG values lower than zero are favourable. The more negative ΔG, the more favourable is the formation of a dimerisation interface.

Under the notion that loop 1 from A3A may influence dimerisation of A3B$_{CTD}$-AL1 (discussed earlier in section 6.3.2.1), we compared the dimerisation potential of our A3B$_{CTD}$-AL1 (A3B$_{CTD}$-QM-ΔL3-AL1swap) and A3B$_{CTD}$-QM-ΔL3 variants using PRISM 2.0 webserver [9, 264] as they only differed by a change of loop 1. We found that the generated A3B$_{CTD}$-AL1 interface models had much lower ΔG values (ΔG ~ -38 and -18 kcal/mol, see Table 8.8 Table 8.9) in comparison to the interface models of our A3B$_{CTD}$-QM-ΔL3 variant (ΔG ~ -13 and -3 kcal/mol, Table 8.6 Table 8.7). This indicated that our A3B$_{CTD}$-AL1 construct had a much greater dimerisation potential than our A3B$_{CTD}$-QM-ΔL3 construct. As A3B$_{CTD}$-AL1 only differed by loop 1 from A3A from our A3B$_{CTD}$-

QM-ΔL3 construct, we can conclude that the presence of the A3A loop 1 directly influences dimerisation. The PRISM 2.0 results are in line with experimental observations, where only the $A3B_{CTD}$-AL1 variant resulted in the elution of a larger putative dimeric species (Figure 6.7).

Two potential dimerisation models of $A3B_{CTD}$-AL1 (termed model 1 and model 2) were predicted. The interface of each model was predicted to be largely formed through electrostatic interactions that accordingly stabilise the structural arrangement of a dimeric molecule (refer to Table 8.8 and Table 8.9). The interface of model 2 was noticeably less favourable (ΔG ~ -18 kcal/mol) than that of the model 1 interface (ΔG ~ -38 kcal/mol) and was predicted to associate through loops 3 and 5, which forms part of the active-site cavity (described in section 1.6.4). As the surface of the active site is highly positively charged due to a histidine and a stretch of arginines, the interactions between residues forming this interface maybe weaker and less stable due to electrostatic repulsion (Figure 6.9B). On the other hand, the interface of model 1 was predicted to be more favourable (ΔG ~ -38 kcal/mol), occurring through interactions slightly away from the active site (Figure 6.9A). Self-association between residues in β1, α1, α2, and loop 1 of the two molecules influences the arrangement of the interface (Figure 6.9). Moreover, this interface places in close proximity a pair of cysteines (Cys239 on beta sheet 2 (β2)) that potentially further stabilise the interface by forming a disulfide bond (see Table 8.8 and Figure 6.9A).

Assessment of these models using an alternate webserver PISA [131-132] indicated that both of these interfaces form through weak interactions such as hydrogen bonds and salt bridges ($\Delta G < 0$) (see Table 8.10 and Table 8.11). So, the predicted disulfide bond of model 1 is essential to structural stability. These results indicate that the interface of model 1 is more likely than that of model 2 interface to form the $A3B_{CTD}$-AL1 dimer, but this remains tentative and needs further verification, for example through SAXS measurements.

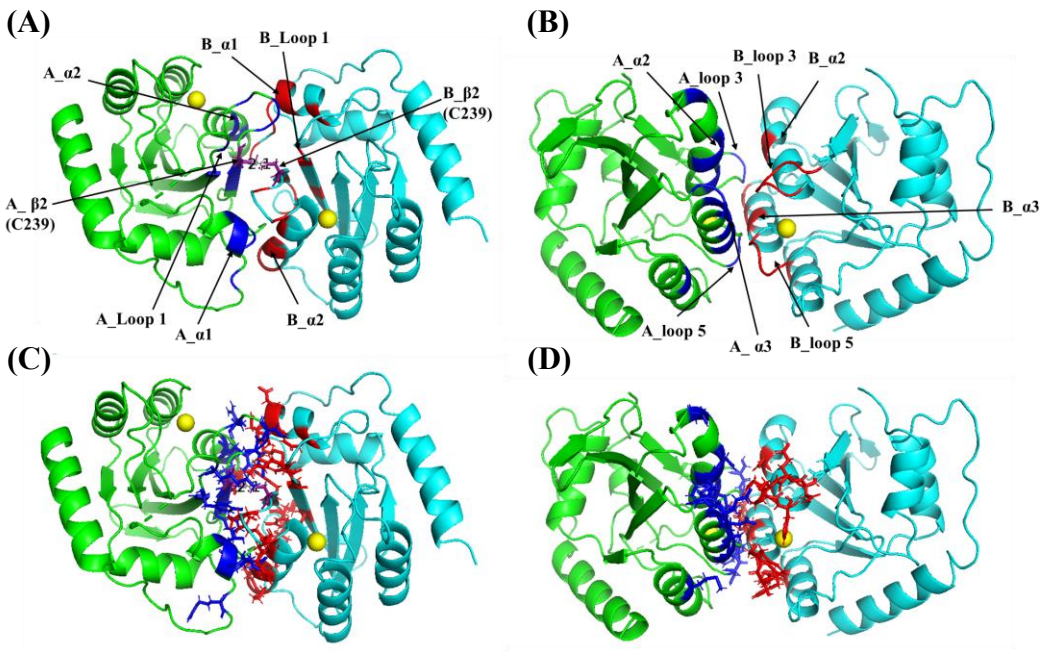


Figure 6.9 Models of the dimerisation interface of A3B$_{CTD}$-AL1.

Predicted interface models between two A3B$_{CTD}$-AL1 molecules using PRISM webserver. A3B$_{CTD}$-AL1 molecule A (green, dimerisation interface in blue) and A3B$_{CTD}$-AL1 molecule B (cyan, dimerisation interface in red) with Zn^{2+} (yellow sphere) in active-site. (A) & (C) Model 1 interface with a calculated ΔG ~ -38.12 kcal/mol, (B) & (D) Model 2 interface with a calculated ΔG ~ -18.6 kcal/mol. Models represented as cartoons in (A) & (B) have the chains and loops involved in the dimerisation interface labelled;, in (C) & (D) residues (in stick format) involved in the dimerisation interface are shown.

Moreover, the A3B$_{CTD}$-AL1 interface models formed interactions between residues within α2 and loop 1 in model 1 (Table 8.8 and Figure 6.9A) or α2, loop 3, and loop 5 in model 2 (Table 8.9 and Figure 6.9B). As discussed in section 1.6.4, several residues in the α2-helix and the aforementioned loops are essential in facilitating substrate binding and deamination of the target dC. Therefore, dimerisation through these regions may hinder substrate accessibility into the active site, consistent with a reduction of dC deamination in comparison to the monomeric A3B$_{CTD}$-AL1 observed in Figure 8.2 (further discussed in section 6.3.4).

6.3.3 SEC-SAXS experiments on the A3B$_{CTD}$-AL1 putative dimer

To examine the hydrodynamic size and relative structure of the A3B$_{CTD}$-AL1 putative dimer, SEC-co-flow-SAXS experiments using 2.5 mg/mL of this protein in kinetic pH 5.5 buffer were performed (slight aggregates formed and the SEC diluted the concentration, so the new calculated concentration based on the absorbance ~ 1.3 mg/mL). Initially, the *R*g trace was overlaid onto the SAXS profile plot indicating that

across the sample peak an approximate *R*g ~ 24 Å can be estimated (frames 540 – 615, Figure 6.10A). This revealed that this protein was in fact larger than the A3B$_{CTD}$-AL1 monomer (*R*g ~18 Å, Figure 6.5). Guinier distribution analysis of the scattering data showed signs of slight sample aggregation at very low scattering angles ($q < 0.0165$ Å^{-1}). Therefore, minor ameliorations were conducted (removing the first few points) to derive a linear dependency of the low *q* data in the Guinier plot (Figure 6.10B insert).

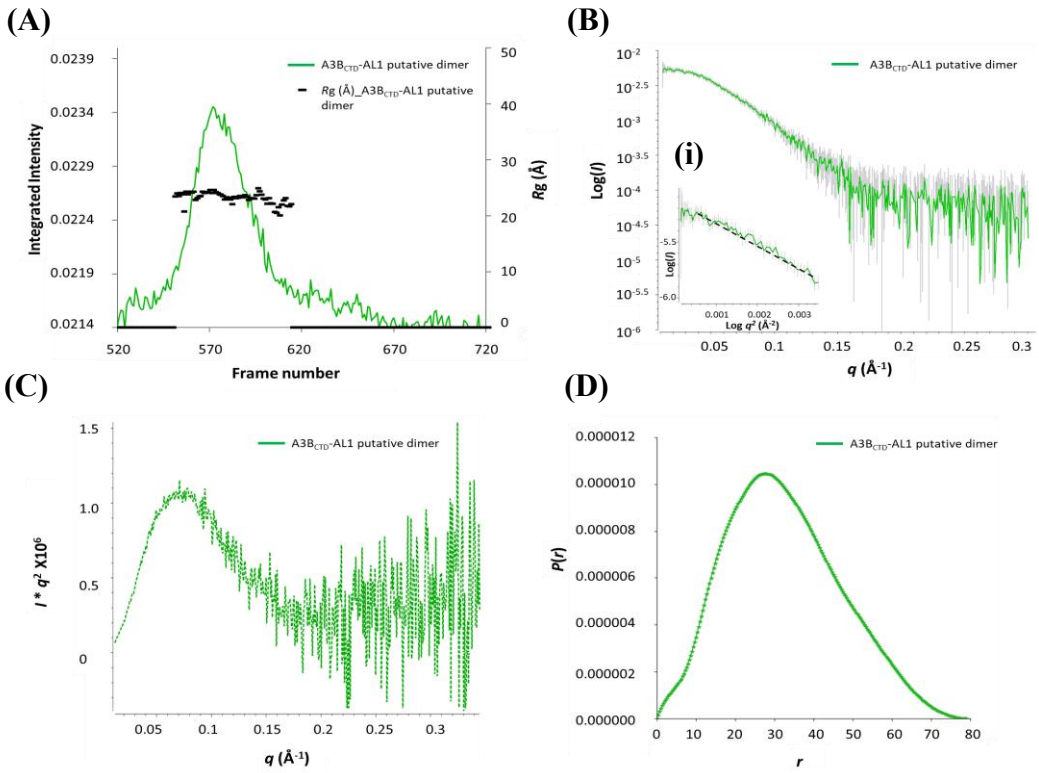


Figure 6.10 SEC-SAXS analysis of A3B$_{CTD}$-AL1 putative dimer.

SAXS analysis of the A3B$_{CTD}$-AL1 putative dimer scattering data was conducted using several programs from PRIMUSQT ATSAS 2.8.3 suite [75]. (A) SAXS profile plot, (B) scattering curve (i) with the Guinier plot insert, (C) Kratky plot, (D) *P*(*r*) distribution plot. Experiments were conducted using 2.5 mg/mL of A3B$_{CTD}$-AL1 putative dimer in kinetic pH 5.5 buffer.

The Kratky plot showed that the sample was a well-folded globular protein. This plot displayed a bell-shaped curve at low *q*, but at higher scattering angles ($q > 0.2$ Å^{-1}) we see an upward turn of data implicating that there may be some flexibility in the putative dimer. Interestingly, the peak maximum of Kratky plot of the A3B$_{CTD}$-AL1 putative dimer was around *q* ~ 0.07 Å, see Figure 6.10C, in contrast the monomeric A3B$_{CTD}$-AL1 Kratky plot that had a peak maximum around *q* ~ 0.10 Å as displayed in Figure 6.5C, further implying that the protein being studied here was hydrodynamically larger in size. This observation was also consistent for the *P*(*r*) plot, which displayed a histogram curve

that was slightly skewed (Figure 6.5D) consistent with standard compact dimers. Both the Guinier and *P*(*r*) calculations resulted in very similar *R*g and *I*(0) values, listed in Table 6.3. The *R*g, D_{max}, and Porod volume of the A3B$_{CTD}$-AL1 putative dimer were also indicative of a bigger molecule (Table 6.3) as compared to the structural parameters derived for the monomeric A3B$_{CTD}$-AL1 (Table 6.1 and Figure 6.5). Finally, calculation of the MW using the Porod volume yields a MW of ~44 kDa similar to the MW derived from *I*(0) of ~46 kDa. This MW was roughly twice the size of the monomeric A3B$_{CTD}$-AL1, verifying that the species eluting in peak 1 observed in Figure 6.7A was in fact a dimeric form of A3B$_{CTD}$-AL1.

Table 6.3 SAXS structural parameters of the A3B$_{CTD}$-AL1 putative dimer

Structural parameters	**A3B$_{CTD}$-AL1 putative dimer**
Guinier analysis	
I(0) (cm^{-1})	0.00480 ± 0.00004
*R*g (Å)	24.2 ± 0.8
*q*min ($Å^{-1}$)	0.0165
q·*R*g max	1.3
Coefficient of correlation (R^2)	0.96
P(r) analysis	
I(0) (cm^{-1})	0.00480 ± 0.00001
*R*g (Å)	24.5 ± 0.6
D_{max} (Å)	77.9
q range ($Å^{-1}$)	0.0165 - 0.3450
Quality estimate	0.88
Porod volume ($Å^3$)	72,900
MW (Porod Volume*0.6) (Da) (ratio to expected subunit 22,000 Da)	43,740 (1.99)
MW (from *I*(0) [178]) (Da) (ratio to expected subunit 22,000 Da)	45,810 (2.08)

Next, space-filled envelope models were generated from the scattering profile (DAMMIF) using *P*2 symmetry as the state of the protein was a dimer. These models were averaged resulting in a model with an acceptable NSD score of 0.816 (Table 6.4). The generated envelope model were first fitted with two A3B$_{CTD}$-AL1 monomer crystal structures (PDB: 5CQI [240]). However, the relative orientations of the two A3B$_{CTD}$-AL1 molecules were difficult to determine.

Therefore, the simulated interface models described earlier in Figure 6.9 were used. The interface models were compared to the A3B$_{CTD}$-AL1 dimer scattering data using rigid body modelling (FoXS [74, 226] and CRYSOL [75]) as described in section 6.1.1. We found that model 1 resulted in very good visual fits to the data (see Figure 6.11A); in contrast model 2 had slightly less optimal visual fits, especially in the mid-*q* region 0.1 – 0.2 Å^{-1} (Figure 6.11B). This suggested that model 1 was the most plausible solution. Furthermore, the interface model 1 was predicted to form in the most favourable conformation (as indicated by ΔG ~ -38 kcal/mol, Figure 6.9A), which further supports these observed results. Chi^2 values for both models did not discriminate, as the Chi^2 values are dominated by the exponentially more intense low *q* <0.1 Å^{-1} data (Figure 6.11A and B).

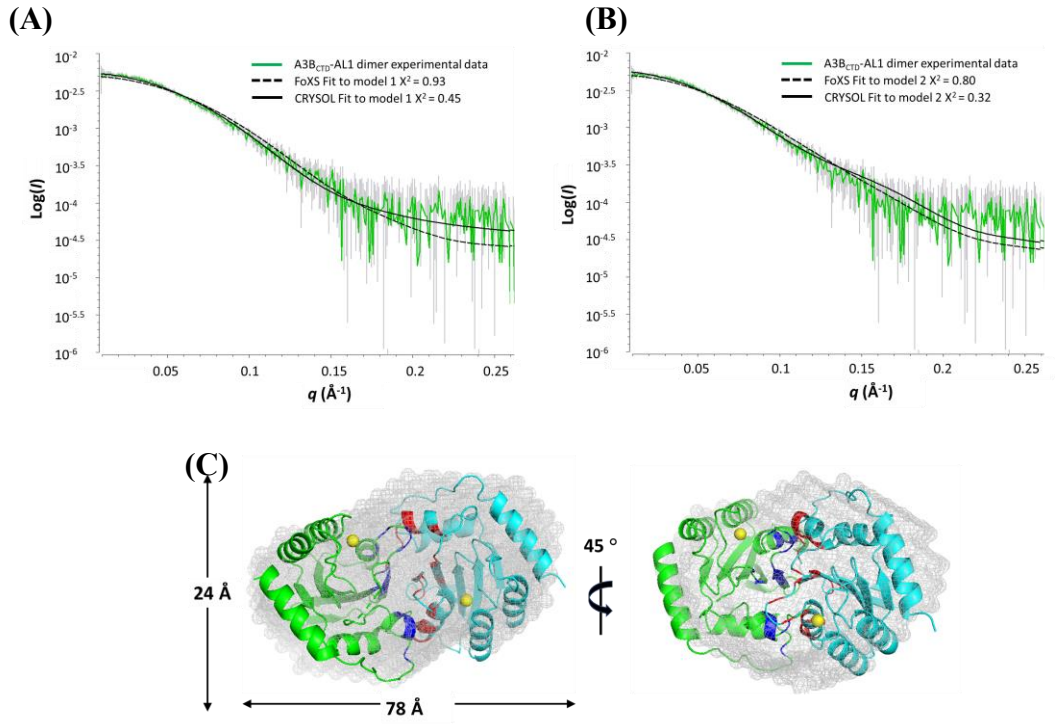


Figure 6.11 Fitting A3B$_{CTD}$-AL1 dimer scattering data to the simulated atomic interface models. Rigid body modelling of the A3B$_{CTD}$-AL1 dimer to (A) interface model 1 or (B) interface model 2 using FoXS and CRYSOL . (C) model 1 interface structure fitted to the A3B$_{CTD}$-AL1 dimer envelope. Averaged envelope model of A3B$_{CTD}$-AL1 dimer derived using *P*2 symmetry (DAMMIF, DAMAVER and DAMFILT, ATSAS 2.8.3 suite [75]) superimposed with interface model 1.

Table 6.4 SAXS fitting and modelling parameters of the A3B$_{CTD}$-AL1 dimer

Modelling parameters	A3B$_{CTD}$-AL1 dimer
Ab intio restoration	
DAMMIF (default parameters, 10 calculations)	
q range (Å^{-1}) for fitting	0.00165 - 0.345
Symmetry	*P*2
NSD (standard deviation)	0.816 (0.110)
Resolution (from SASRES) (Å)	41 ± 3
Structure modelling	
Interface dimer model 1	
FoXS, Chi^2	0.93
CRYSOL, Chi^2	0.45
Interface dimer model 2	
FoXS, Chi^2	0.80
CRYSOL, Chi^2	0.32

Moreover, model 1 interface was found to be organised in a similar manner to a previously reported crystal structure of an A3A homodimer (PBD 4XXO [23]). In this structure a dimerisation groove was formed through the N-terminus slightly away from the active-site, leading to binding of the two A3 molecules [23], in a manner similar to our predicted interface model 1 (Figure 6.12B). Formation of this interface between the two A3A molecules were reported to occur through interactions with metal ions, salt bridges, and hydrophobic interactions. Four main residues (His11, Lys30, His16, and His56) were identified to be necessary for the formation of this interface (displayed in Figure 6.12A). One important interaction in the A3A homodimer was between a pair of His56 that were bridged by a Zn^{2+} ion, interestingly the A3A His56 is isosteric with Cys239 of A3B$_{CTD}$-AL1. Furthermore, interactions occurred between residues in loop 1, α1, α2, and β2 (Figure 6.12A), which were equivalent to the interactions of the dimeric A3B$_{CTD}$-AL1.

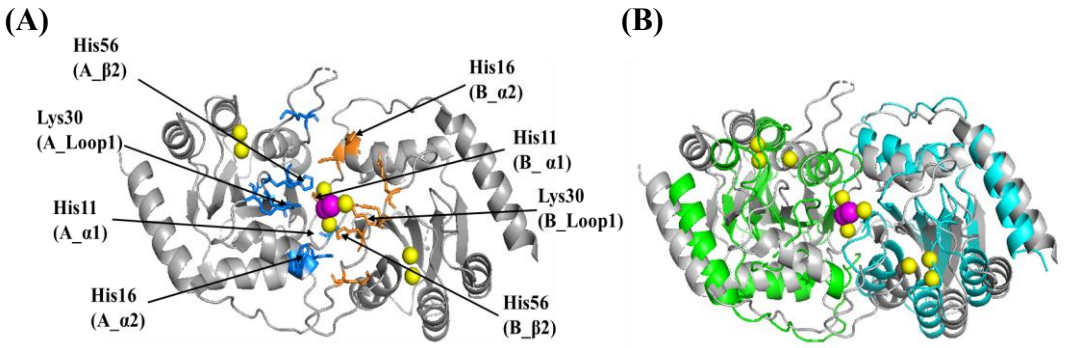


Figure 6.12 Dimerisation interface of A3A and A3B$_{CTD}$-AL1.

(A) A3A homodimer (PDB 4xxo [23]) (grey) labelled with residues that are important in forming the dimerisation interface. A3A molecule A dimerisation residues labelled in blue and A3A molecule B dimerisation residues labelled in orange (B) Comparison of the A3A homodimer (PDB 4xxo [23]) (grey) to our dimer interface model 1 of A3B$_{CTD}$-AL1 (molecule A in green and molecule B in cyan). Zn^{2+} ions are represented as yellow spheres and chloride ions are represented as magenta spheres.

Therefore, from this we can presume that our derived in-solution A3B$_{CTD}$-AL1 dimer may in fact resemble the model 1 interface. Superimposing the generated envelope model with our interface dimer model 1 structure we can see that it fits well (Figure 6.11A and C). This model gives a starting point to look at the association of subunits of two-domain A3 enzymes.

6.3.4 A3B$_{CTD}$-AL1 dimer complex with ssDNA

The A3B$_{CTD}$-AL1 dimeric form was identified to be catalytically active, with behaviour intermediate between the A3B$_{CTD}$-QM-ΔL3-AL1swap and A3B$_{CTD}$-QM-ΔL3 (see Figure 8.2). Therefore, to understand how ssDNA binds to the A3B$_{CTD}$-AL1 dimer in order to permit catalysis, the relative structure of this dimer in complex with dZ-oligo (to prevent complications of activity) was explored using SEC-co-flow-SAXS. Data were collected for the A3B$_{CTD}$-AL1 dimer at 2.5 mg/mL in kinetic pH 5.5 buffer with and without dZ-oligo (slight aggregates formed and the SEC diluted the concentration, adjusted ~ 1.3 mg/mL as described in section 6.2.1). However, in these experiments we used a higher concentration of the dZ-oligo at a 1 to 10 ratio (protein to oligo) to ensure that the A3B$_{CTD}$-AL1 dimer was saturated with ssDNA and drive the formation of a protein-ssDNA complex. The initial inspection of the UV elution profile (Figure 6.13C) and the SAXS profile plot (Figure 6.13B) showed that the *R*g values were not constant across the profiles, signalling that the sample was not homogeneous. Only the peak corresponding to the dZ-oligo had a constant *R*g value of ~10 Å, as expected (711-750 sec, Figure 6.13B & C, frame numbers 668 – 698 Figure 6.14).

Next, we compared the A3B$_{CTD}$-AL1 dimer/dZ-oligo elution and SAXS profiles (Figure 6.13C & Figure 6.14A) to the profiles of the ligand-free A3B$_{CTD}$-AL1 dimer (Figure 6.13A & Figure 6.14A). We found that upon the presence of the dZ-oligo the intensity of the peak corresponding to the A3B$_{CTD}$-AL1 dimer decreases significantly (elution profile 550 – 640 sec, Figure 6.13C, SAXS profile frames 520 – 610, Figure 6.14A). Interestingly, upon the decrease of the A3B$_{CTD}$-AL1 dimer peak we noticed the appearance of a peak coinciding with the relative elution position and size (Rg ~ 18-19 Å) of the monomeric A3B$_{CTD}$-AL1 species (elution profile shoulder peak 650 – 700 sec insert Figure 6.13C, and SAXS profile frames ~ 613 – 653 Figure 6.14A).

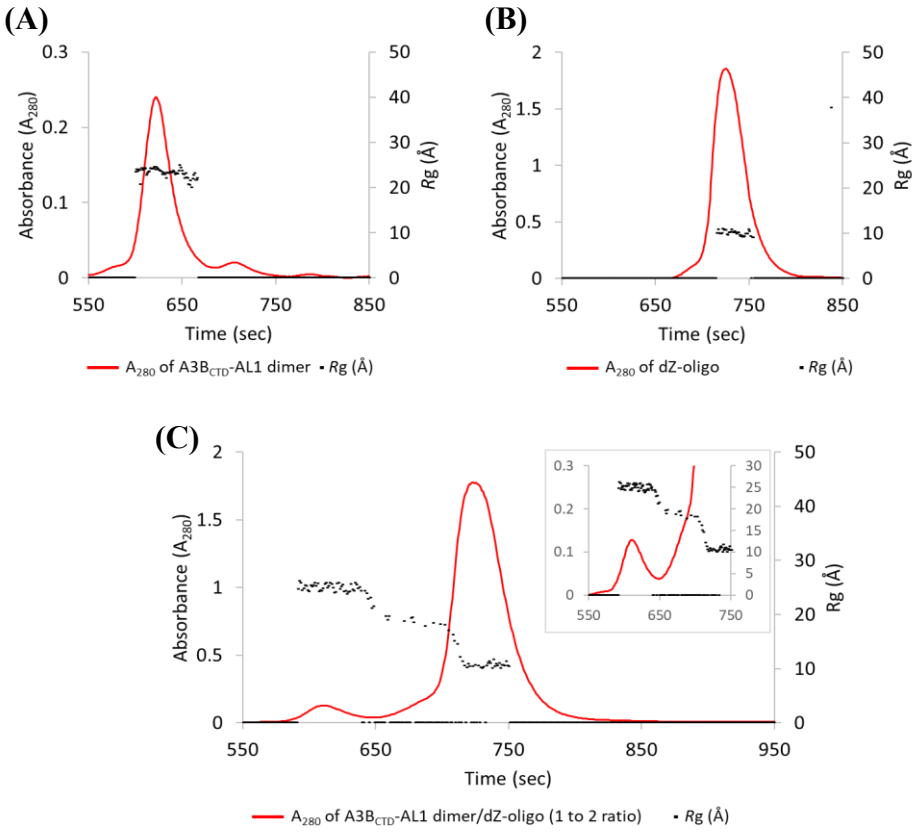


Figure 6.13 SEC elution profiles of A3B$_{CTD}$-AL1 dimer and A3B$_{CTD}$-AL1 dimer/dZ-oligo.
SEC-FPLC performed using (A) 2.5 mg/mL (~ 56 µM) of A3B$_{CTD}$-AL1 dimer, (B) 1.7 mg/mL (~ 566 µM) of dZ-oligo, and (C) A3B$_{CTD}$-AL1 dimer with dZ-oligo at a 1 to 10 ratio, closeup shown in the insert. Samples were run at 25 °C through a SEC column at a flow rate of 0.2 mL/min in kinetic pH 5.5 buffer (50 mM citrate-phosphate pH 5.5, 0.2 M NaCl, 2 mM β-ME).

The increase of this monomeric species was intriguing. The three peaks visualised in Figure 6.14B reflected that the A3B$_{CTD}$-AL1/dZ-oligo sample contained three scattering species, which were postulated to be a A3B$_{CTD}$-AL1 dimer species, an A3B$_{CTD}$-AL1 monomeric species, and the unbound dZ-oligo (termed species 1-3, respectively). From

the UV elution profile we observed that the peak maximum associated with the unbound dZ-oligo decreases slightly by ~5-fold (Figure 6.13C) in comparison to the control oligonucleotide elution profile in Figure 6.13B. This decrease suggested that a protein-ssDNA complex may have formed. However, it was unclear from the elution profile which A3B$_{CTD}$-AL1 species was in complex with the ssDNA (Figure 6.13C). To analyse each species the scattering curves had to first be generated, but as components of each peak overlapped slightly the boundaries of each scattering species had to be defined using deconvolution analysis SVD/EFA (BioXTAS RAW [105] (described in section 6.1.1 and Figure 8.9). SVD analysis was first performed resulting in an eigenvalue of three (see Figure 8.9), corresponding to the presence of three distinct scattering species in the sample, as previously presumed from the *R*g pattern in the elution and SAXS profiles (Figure 6.13C and Figure 6.14A). Then, EFA was used to define the boundaries of each species (Figure 6.14B) and their respective 1D-scattering curves were generated.

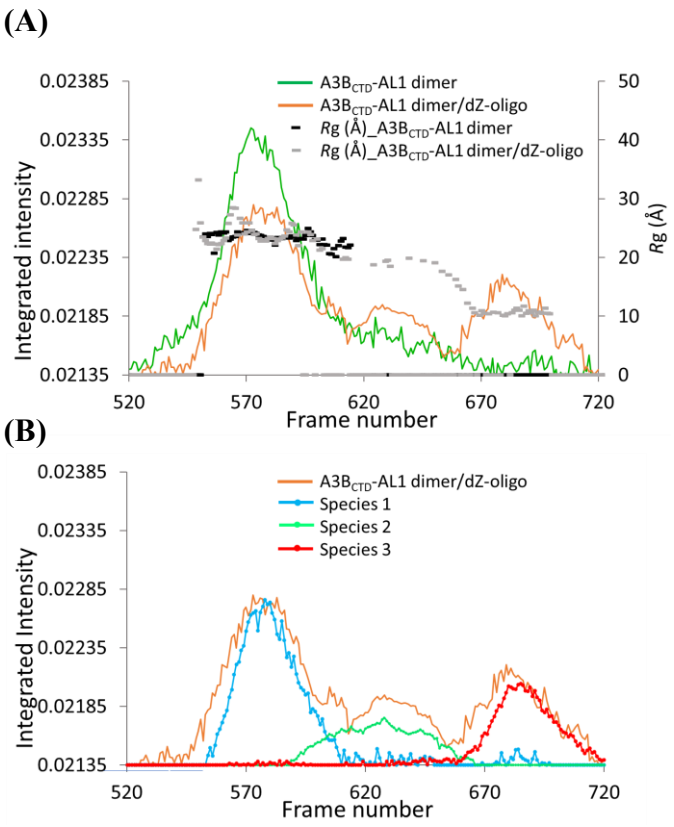


Figure 6.14 Deconvolution of SAXS data of the A3B$_{CTD}$-AL1 dimer/dZ-oligo.

(A) Comparison of the SAXS profiles of the A3B$_{CTD}$-AL1 dimer (green) and A3B$_{CTD}$-AL1 dimer with dZ-oligo at a 1 to 10 ratio (orange) overlaid with *R*g trace values. (B) EFA analysis using BioXTAS RAW indicating the distinct scattering species 1-3.

Analysis of these scattering curves was performed as described in section 6.1.1 using the PRIMUSQT ATSAS 2.8.3 suite. The scattering curve of species 3, which corresponds to the unbound dZ-oligo was not further analysed due to limitations of SAXS resolution (> 10 Å). So, focus was placed on the analysis of species 1 and 2. First, the double log plot indicated that both species were homogeneous as the data at low *q* plateaued (Figure 6.15B). The Kratky plots reinforced that species 1 was larger than species 2, and based on the peak maximum position we can justify that species 1 was fundamentally the dimeric A3B$_{CTD}$-AL1 (q ~ 0.07 Å^{-1}), while species 2 was the A3B$_{CTD}$-AL1 monomer (q ~ 0.1 Å^{-1}) (Figure 6.15C) as observed earlier in sections 6.2.3 and 6.3.3), and supported by the calculated MW (see Table 6.5)

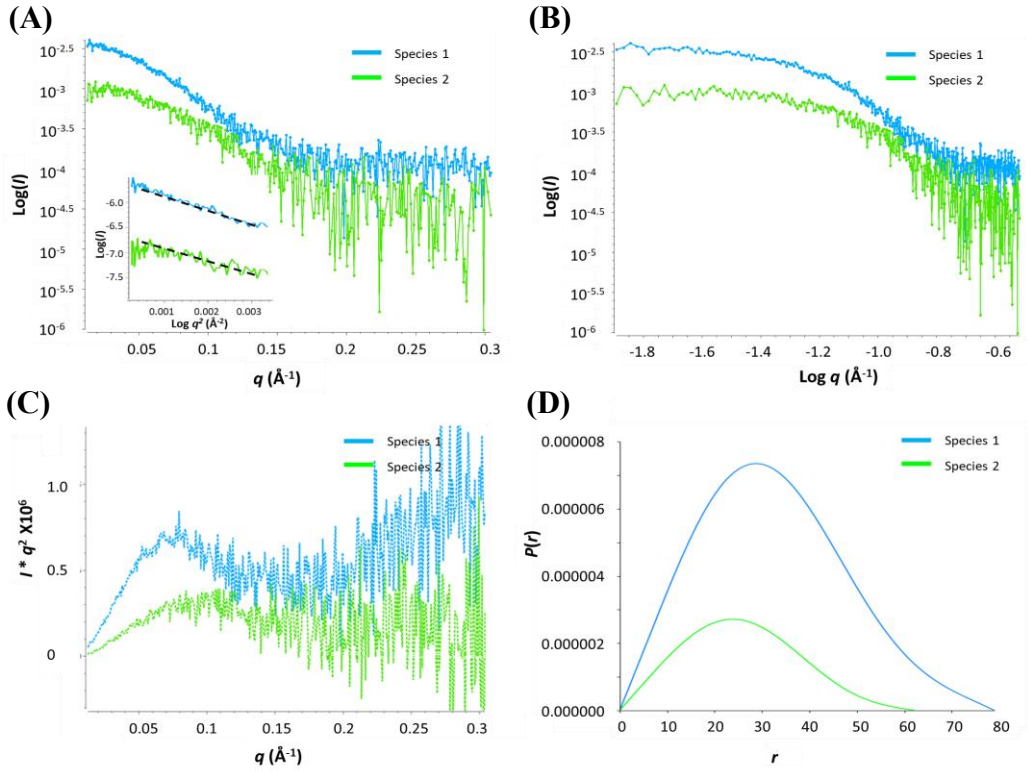


Figure 6.15 SAXS analysis of the A3B$_{CTD}$-AL1-dimer/dZ-oligo deconvoluted species.
Analysis of the scattering data of the deconvoluted A3B$_{CTD}$-AL1 dimer/dZ-oligo species was conducted using several programs from the PRIMUSQT ATSAS 2.8.3 suite [75]. (A) 1D-scattering curves of each species with a Guinier plot insert, (B) double log plot, (C) Kratky plot, (D) *P*(*r*) distribution plot. Experiments were conducted using 2.5 mg/mL of A3B$_{CTD}$-AL1-dimer with dZ-oligo at a 1 to 10 ratio in kinetic pH 5.5 buffer.

Table 6.5 SAXS structural parameters of the A3B$_{CTD}$-AL1 dimer/dZ-oligo sample

Structural parameters	Species 1 (A3B$_{CTD}$-AL1 dimer)	Species 2 (A3B$_{CTD}$-AL1 monomer/dZ-oligo)
Guinier analysis		
$I(0)$ (cm^{-1})	0.0036 ± 0.0001	0.00110 ± 0.00005
Rg (Å)	24.8 ± 1.5	20.0 ± 1.5
$q$min (Å^{-1})	0.01795	0.0231
$q \cdot R$g max	1.3	1.3
Coefficient of correlation (R^2)	0.98	0.86
P(r) analysis		
$I(0)$ (cm^{-1})	0.00352 ± 0.00007	0.00110 ± 0.00005
Rg (Å)	25.1 ± 0.5	20.1 ± 0.8
D_{max} (Å)	78.2	60.1
q range (Å^{-1})	0.01795 - 0.2756	0.0231 - 0.2317
Quality estimate	0.89	0.85
Porod volume (Å^3)	74,000	38,400
MW (Porod Volume*0.6) (Da) (ratio to expected subunit 22,000 Da)	44,400 (2.02)	23,040 (1.05)

The molecular weights of these species were only calculated from the Porod volume, not from $I(0)$ [178] (listed in Table 6.5), as the concentration of each species was difficult estimate due to their overlap during elution. The Guinier distribution analysis showed that the scattering profile of species 1 had a reasonably good linear fit, after eliminating the first several lowest q data points (Figure 6.15A insert), with estimated Rg and $I(0)$ values (listed in Table 6.5) in good agreement with values calculated from the $P(r)$ plot (as described in section 6.3.3). However, analysis performed using the low q data range for species 2 required significant amendments (removing the first 15 points), giving a suboptimal fit to the linear regression, with compromised precision and accuracy (Figure 6.15A insert). The $P(r)$ plot for species 2 gave Rg ~ 20 Å and D_{max} ~ 60 Å and the Porod plot gave MW ~ 23 kDa (listed in Table 6.5). These parameters are consistent with a monomeric A3B$_{CTD}$-AL1 in complex with ssDNA (as noted earlier in Table 6.1). However, as the scattering data of species 2 were particularly noisy (Figure 6.15A) due to its low concentration (see Figure 6.13C and Figure 6.14), calculation of its structural parameters are less precise than those for species 1 (see Table 6.5)

The values of *R*g, D_{max}, and MW (from Porod volume) of species 1 (listed in Table 6.5) are similar to those obtained for the ligand-free A3B$_{CTD}$-AL1 dimer (Table 6.3). This suggested that in the presence of the oligonucleotide the A3B$_{CTD}$-AL1 dimer did not form a complex with the ssDNA. To verify this notion, the A_{260}/A_{280} ratio of the ligand-free A3B$_{CTD}$-AL1 dimer was compared to the ratio of the A3B$_{CTD}$-AL1 dimer/dZ-oligo sample to assess the presence of DNA in the eluted protein sample. Using this method, we observed that the ratio remained relatively unchanged in the region where species 1 elutes off the SEC column (see Figure 6.16). This implied that the A3B$_{CTD}$-AL1 dimer (species 1) was in fact not bound to the oligonucleotide. Interestingly, as shown in Figure 6.16, in the region where species 2 elutes the A_{260}/A_{280} ratio of the A3B$_{CTD}$-AL1 dimer/dZ-oligo sample (1.335 at 680 s) is significantly larger than the ratio of the ligand-free A3B$_{CTD}$-dimer sample (0.811 at 680 s).

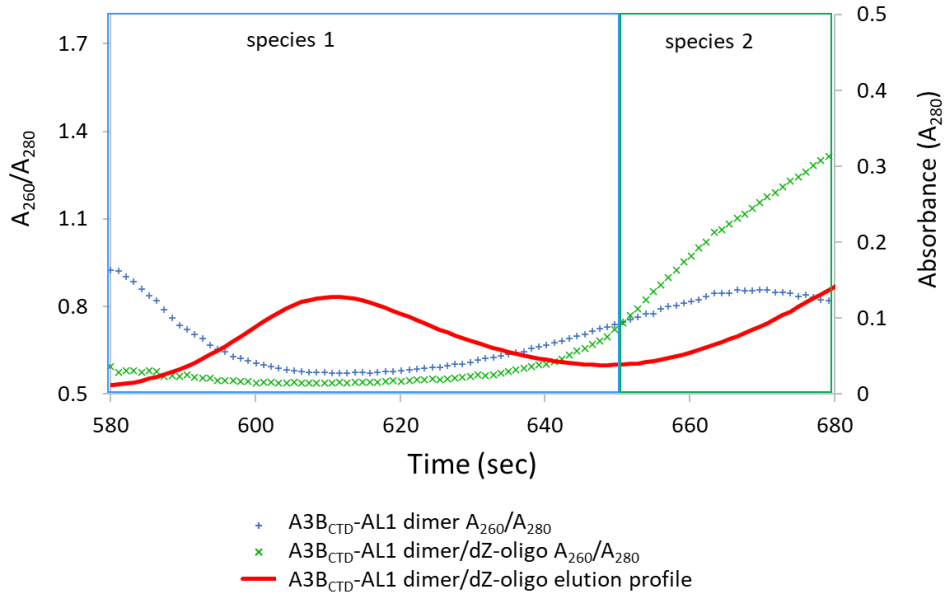


Figure 6.16 Ratio of A_{260}/A_{280} to assess the presence of DNA during elution of A3B$_{CTD}$-AL1 dimer/dZ-oligo sample.

A_{260}/A_{280} ratio taken of the A3B$_{CTD}$-AL1 dimer (blue) and the A3B$_{CTD}$-AL1 dimer/dZ-oligo (green) samples. Overlaid with the A_{280} elution profile of the A3B$_{CTD}$-AL1 dimer/dZ-oligo sample (red) to display the boundaries of each species (species 1, and 2).

These results justify earlier observations that this monomeric A3B$_{CTD}$-AL1 is in complex with the dZ-oligo. As our dimeric A3B$_{CTD}$-AL1 (species 1) does not bind ssDNA in solution we can presume that it is not associated with the observed catalytic activity (Figure 8.2). So, we can conclude that the trace of monomeric A3B$_{CTD}$-AL1 (species 2) that is always present would be mediating the observed activity. From these observations, we can propose that the catalytic activity of the enzyme could be regulated through

dimerisation as suggested by earlier studies [22-23, 27, 79, 118, 233]. However, whether dimerisation of A3 enzymes would occur in this manner within the cell remains unknown.

Table 6.6 SAXS fitting and modelling parameters of the A3B$_{CTD}$-AL1 dimer/dZ-oligo sample.

Modelling parameters	**Species 1 (A3B$_{CTD}$-AL1 dimer)**	**Species 2 (A3B$_{CTD}$-AL1 monomer/dZ-oligo)**
Ab intio restoration		
DAMMIF (default parameters, 10 calculations)		
q range (Å^{-1}) for fitting	0.01795 - 0.2756	0.0231 - 0.2317
Symmetry	P2	P1
NSD (standard deviation)	0.993 (0.081)	1.006 (0.044)
Resolution (from SASRES) (Å)	28 ± 2	29 ± 2
Structure modelling		
5TD5*		
FoXS, Chi^2	2.46	0.30
CRYSOL, Chi^2	1.73	0.28
Interface dimer model 1		
FoXS, Chi^2	1.03	0.42
CRYSOL, Chi^2	1.41	0.33
Interface dimer model 2		
FoXS, Chi^2	1.26	0.42
CRYSOL, Chi^2	0.77	0.33

Next, space-filled envelope models were produced from the scattering profiles of species 1 using *P*2 symmetry and species 2 using *P*1 symmetry (DAMMIF). However, as expected the high noise ratio of the data generated envelope models that were largely variable resulting in high NSD scores > 0.9 for the averaged models (for satisfactory models NSD ≤ 0.8); see Table 6.6. Therefore, the derived envelope models do not give an accurate representation of the SAXS model of each species (see Figure 8.11).

So, rigid body modelling was performed (FoXS and CRYSOL) to provide better insights into the structures of each species by comparing their experimentally derived scattering data to the simulated dimer models described earlier (section 6.3.2.2 and 6.3.3) and the atomic A3B$_{CTD}$/dZ-oligo structure (5TD5* mentioned earlier 6.2.3 [239]) (Table 6.6). Using this method, we found that monomeric A3B$_{CTD}$-AL1/dZ-oligo (species 2) scattering data had a nice visual fit to the A3B$_{CTD}$-ssDNA structure (5TD5*), particularly in low *q* range of 0-0.15 Å^{-1} (Figure 6.17B i). In contrast, this species had less then optimal visual fits to the A3B$_{CTD}$-AL1 dimer interface models (Figure 6.17B ii-iii), as anticipated. However, the Chi^2 values did not differentiate between the models as Chi^2 is dominated by the exponentially more intense low *q* data (< 0.1 Å^{-1}). On the other hand, our A3B$_{CTD}$-AL1 dimeric species data (species 1) resulted in a very poor fit to the monomeric A3B$_{CTD}$-ssDNA structure (5TD5*), as expected (see Figure 6.17A i and Table 6.6 which reports high Chi^2 values). Better fits, however, were obtained against the dimer interface models, especially at low *q* regions (0.01-0.12 Å^{-1}, Figure 6.17A ii-iii).

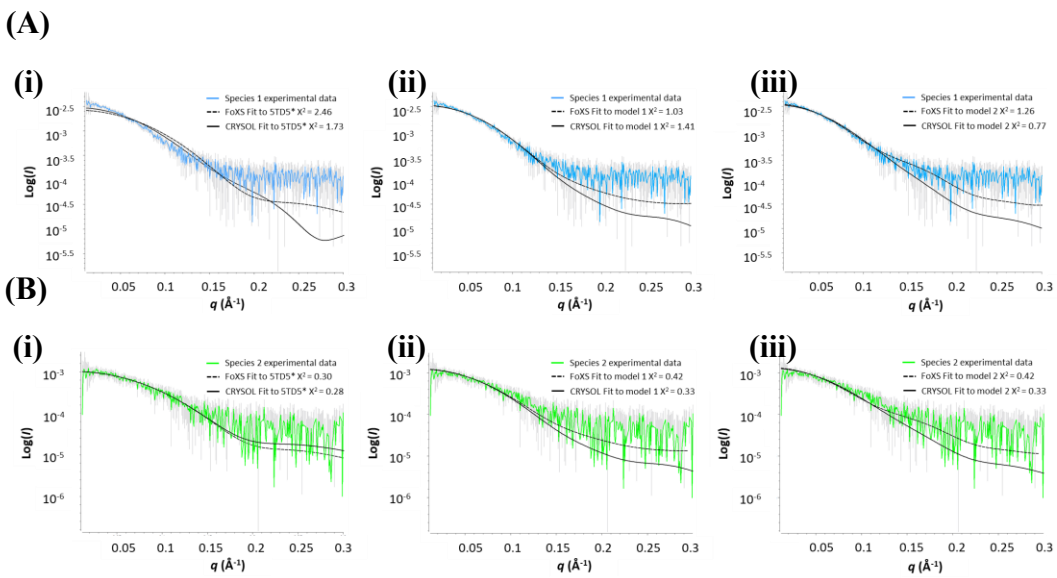


Figure 6.17 Fitting A3B$_{CTD}$-AL1 dimer/dZ-oligo species 1 and 2 scattering data to models structures.

Rigid body modelling of the A3B$_{CTD}$-AL1 dimer/dZ-oligo (A) species 1 and (B) species 2, against the (i) 5TD5* crystal structure (A3B$_{CTD}$-9mer), (ii) model 1 dimer interface (iii) model 2 dimer interface using FoXS and CRYSOL.

However, at $q > 0.15$ Å^{-1}, this dimeric species did not seem to fit so nicely to these models as previously observed for our ligand-free A3B$_{CTD}$-AL1 dimer (Table 6.2). The discrepancy is attributed to errors in deconvoluting noisy data, rather than to some

interaction of the $A3B_{CTD}$-AL1 dimer with the dZ-oligo, as such interaction is not supported by the A_{260}/A_{280} ratio.

6.4 SAXS models of the double domain A3 protein

Structures of inactive single A3 domains in complex with substrate ssDNA have been reported [239], but structures of double-domain A3 bound to ssDNA remain unknown due to difficulties in crystallisation. Therefore, to understand how one domain is positioned relative to the other and how ssDNA binds to these domains in solution, SEC-SAXS was utilised. We examined active two-domain A3 structures in the absence and presence of our ssDNA-based inhibitor to prevent substrate deamination during the experiment.

6.4.1 Selection of the double domain A3G

Initially, A3B constructs, cloned into a pcDNA3.1(-)/MycHis expression vector (Invitrogen) (obtained from Harris and Aihara labs, University of Minnesota), were transfected into HEK293T mammalian cells and purified as described [239-240]. However, expression levels and solubility of the full-length recombinant A3B protein were poor. Therefore, the closely related A3G enzyme was used to model two-domain A3 structures (obtained from Prof. Linda Chelico, University of Saskatchewan). Both A3G and A3B are subcategorised into the same phylogenetic class based on the same domain configuration (N'-Z2/Z1-C') (see section 1.6.1) and share an overall sequence identity of approximately ~56 % (refer to Figure 8.10).

6.4.2 The double domain A3G

A catalytically active full-length wild type A3G protein (termed A3Gfl-Wt) was assessed at 2.5 mg/mL in A3G kinetic pH 6 buffer to favour DNA binding (slight aggregates formed and SEC caused a dilution so a new concentration ~ 1.3 mg/mL was calculated based on the absorbance). Inspection of the UV elution profile and SAXS profile of the A3Gfl-Wt showed that the sample was not well resolved, resulting in a large peak made up of several overlapping peaks (Figure 6.18A and B). Moreover, superimposing the *R*g trace values over the elution and SAXS profiles we see a large variability in the *R*g values from ~30 to !50 Å across this peak, indicating that multiple scattering species were present in the sample (Figure 6.18A and B). Therefore, before continuing to parameter analysis, the SAXS data had to first be analysed by SVD and EFA (SVD/EFA BioXTAS RAW [105]) to elucidate the number and boundaries of the scattering components (see Figure 8.12). This analysis illustrated that there were three potential components present

in the sample (EFA 1; frames 353-578, EFA 2; 463-668, EFA 3; 563-718). However, as the boundaries were significantly overlapped it was difficult to isolate the data of each scattering species (Figure 6.18C).

Therefore, Gaussian decomposition analysis (US-SOMO) was performed by Dr. Timothy Ryan (SAXS/WAXS beamline scientist, Australian Synchrotron) by fitting Gaussian functions to produce decomposed 1D-scattering curves of distinct components (Figure 6.18D). The derived scattering curves had a high noise-to-signal ratio as a consequence of deconvolution and the low concentration of each species. Although the data were fitted with five Gaussians, only Gaussians 2 and 3 (termed A3G_species A and B) produced decent scattering curves (Gaussian 2, frames 414-483, and Gaussian 3, frames 459-541) with higher intensities than the other Gaussians indicating their significance in the overall scattering pattern of the sample. Therefore, these two curves were analysed using PRIMUSQT (ATSAS 2.8.3 suite) to identity the species present.

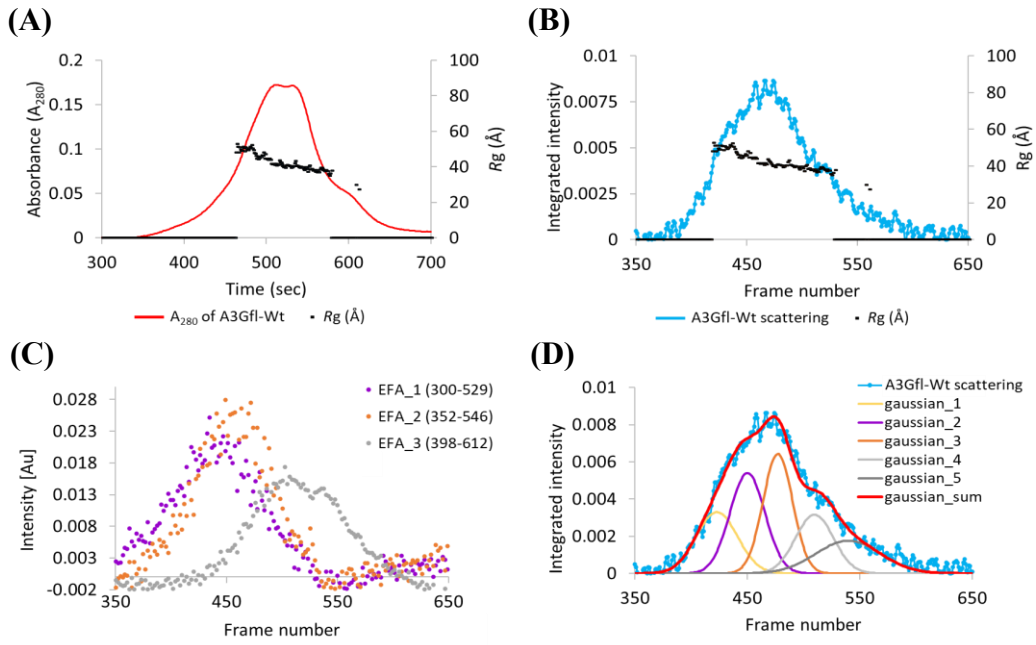


Figure 6.18 Deconvolution of the SAXS data of the A3Gfl-Wt.

SEC-SAXS analysis of A3Gfl-Wt. (A) SEC elution profile collected with A_{280} and (B) SAXS profile superimposed with *R*g trace values. (C) EFA analysis using BioXTAS RAW (D) Gaussian decomposition analysis using US-SOMO. Experiments were conducted at 25 °C using 2.5 mg/mL A3Gfl-Wt in A3G pH 6.0 kinetic buffer (50 mM phosphate pH 6.0, 200 mM NaCl, 2mM β-ME, 5% glycerol, 200 µM Na_2-EDTA).

Initial analysis of the derived scattering curves revealed that the A3G scattering species were homogeneous, illustrated by the plateau of low *q* values in the double log graph (Figure 6.19B). Guinier distribution analysis of A3G_species A using data in the low *q* range resulted in an unsatisfactory fit to the linear regression, which compromised estimation of the *R*g and *I*(0) (Figure 6.19B insert). From the *P*(*r*) plot, A3G_species A was estimated to have an *R*g ~ 60 Å and D_{max} ~ 160 Å. However, as the scattering data of this species had a low signal-to-noise ratio, accurate derivation of the structural parameters for this species was difficult (Figure 6.19A and B).

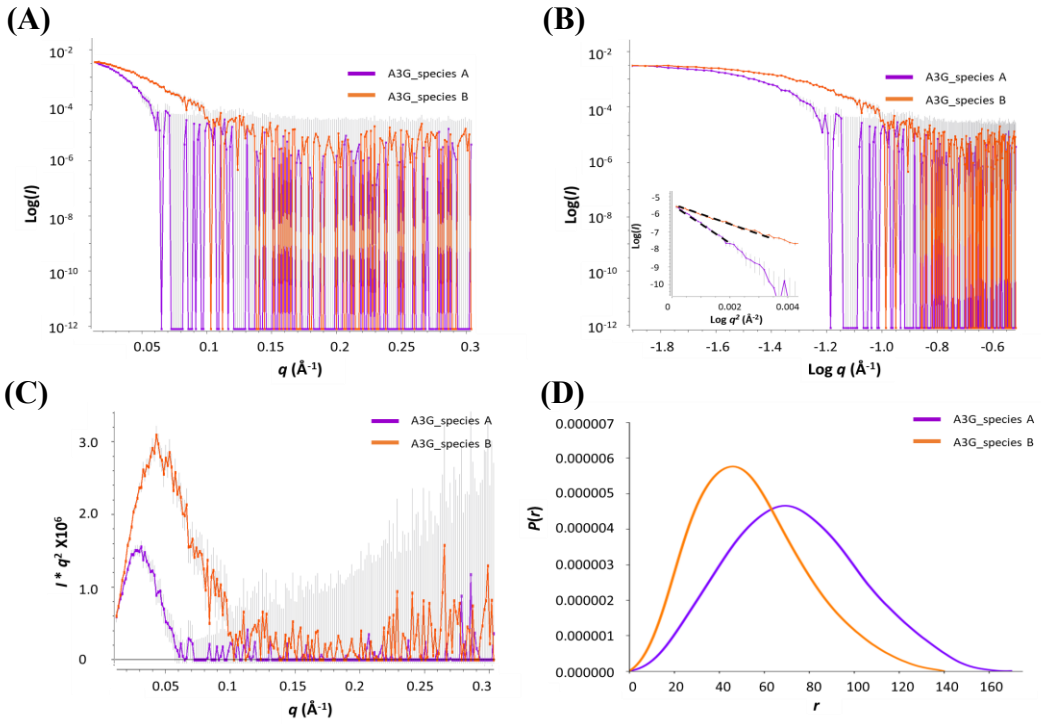


Figure 6.19 SAXS analysis of A3Gfl-Wt deconvoluted species.

Analysis of the scattering data of the deconvoluted A3Gfl-Wt species was conducted using several programs from the PRIMUSQT ATSAS 2.8.3 suite [75]. (A) 1D-scattering curves of the species (B) Double log plot with insert Guinier plot (C) Kratky plot. (D) *P*(*r*) distribution plot. Experiments were conducted using 2.5 mg/mL of A3Gfl-Wt in A3G pH 6.0 kinetic buffer.

Comparison of the Kratky plots of the derived scattering species showed that the position of the peak maximum of the A3G_species A (at *q* ~0.025 $Å^{-1}$) was half that of the A3G_species B (*q* ~ 0.05 $Å^{-1}$) (Figure 6.19C), indicating that the A3G_species A was about double the size of the A3G_species B. This was also observed in the elution pattern, as bigger particles elute off the SEC column earlier than smaller ones, A3G_species A elutes before A3G_species B (Figure 6.18A). Finally, A3G_species B was perceived to

be much more prominent than the larger A3G_species A, as displayed by better signal to noise (Figure 6.18B).

We then focused, on the more prominent A3G_species B, for which we found a sensible fit of the Guinier plot indicating a monodispersed sample (see Figure 6.19B insert). Both the *R*g and *I*(0) derived from the Guinier and the *P*(*r*) plots were very similar, confirming the relative size of the species as listed in Table 6.7. The Kratky plot showed a nice bell-shaped curve with a peak maximum $q \sim 0.05$ Å^{-1}; however, the plot at high q did not completely return to the baseline suggesting that the structure may contain flexible regions (Figure 6.19C). Noticeably, the estimated size of *R*g and D_{max} of our A3G_species B (*R*g ~ 42 Å, D_{max} ~ 141 Å, Table 6.7) was found to be similar to a previously reported A3G elongated dimer SAXS model [276]. However, our molecular weight was estimated to be ~ 208 kDa (derived from Porod volume, not from *I*(0) [178] as each species overlapped discussed in section 0), which was double the size of this previously reported A3G elongated dimer ~100 kDa [276]. This suggested that our A3G species is not a dimer, but may exist as a compact tetramer under our conditions, where four two-domain A3G molecules oligomerise together.

Table 6.7 SAXS structural parameters of the A3Gfl_Wt sample

Structural parameters	**A3G_species B (A3G tetramer)**
Guinier analysis	
I(0) (cm^{-1})	0.0044 ± 0.0001
*R*g (Å)	41.5 ± 1.5
*q*min (Å^{-1})	0.01245
q·*R*g max	1.3
Coefficient of correlation (R^2)	0.95
P(r) analysis	
I(0) (cm^{-1})	0.0044 ± 0.0001
*R*g (Å)	42.0 ± 1.0
D_{max} (Å)	145.9
q range (Å^{-1})	0.01245 - 0.34030
Quality estimate	0.85
Porod volume (Å^3)	346,000
MW (Porod Volume*0.6) (Da) (ratio to expected subunit 46,408 Da)	202,600 (4.4)

Therefore, *ab initio* shape restoration (DAMMIF, ATSAS 2.8.3 suite [75]) was performed with a *P*4 symmetry constraint based on the parameter calculations indicating a tetrameric structure. The averaged low-resolution dummy model had an acceptable NSD value of 0.645 indicating that the models were in good agreement (see Table 6.8). Modelling the molecular envelope without symmetry gave a similar NSD ~ 0.695.

Table 6.8 SAXS fitting and modelling parameters of the A3Gfl-Wt sample

Modelling parameters	**A3G_species B**
Ab initio restoration	
DAMMIF (default parameters, 10 calculations)	
q range ($Å^{-1}$) for fitting	0.01245 - 0.34030
Symmetry	*P*4
NSD (standard deviation)	0.645 (0.180)
Resolution (from SASRES) (Å)	33 ± 3

The envelope model exhibited an elongated shape (Figure 6.20A), consistent with observations from the Kratky plot (Figure 6.19C). Initially, we attempted to superimpose the derived A3G space-filled model with four of our $A3B_{CTD}$-AL1 dimer envelopes to determine their relative fit, resulting in a generally good fit (Figure 6.20B). However, due to the large surface area of the A3G tetramer space-filled model we were not able to accurately determine the position of these envelopes, so we cannot build a model structure using this approach. Nevertheless, docking the four $A3B_{CTD}$-AL1 dimer envelopes onto this A3G tetramer model we saw that they were tightly attached to one another, consistent with notation of being a compact tetramer.

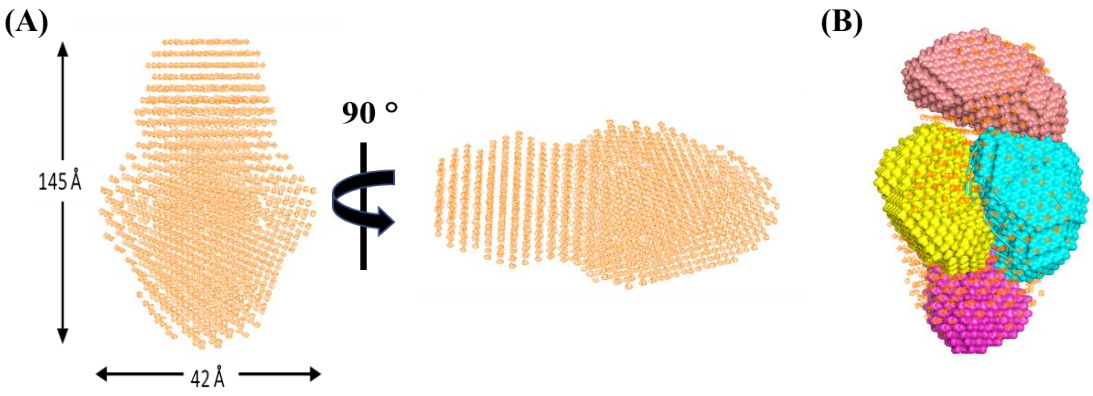


Figure 6.20 *Ab initio* shape restoration of A3G tetramer.

(A) Envelope models generated using DAMMIF *(*ATSAS 2.8.3 suite) [75] under *P*4 symmetry of A3G tetramer and further refined using DAMAVER and DAMFILT (ATSAS 2.8.3 suite) [75], (B) A3G tetramer envelope model (orange) docked with four $A3B_{CTD}$-AL1 dimer envelopes.

Until now, double-domain A3 structures remained unknown, but crystal structures of A3G dimers (derived from the Rhesus macaque) have been published earlier this year (January 2020) (PDB; 6p40, 6p3x, 6p3z, and 6p3y) [286]. Using these atomic A3G dimer structures, interface analysis can be performed using PRISM 2.0 webserver [9, 264] in a similar manner as described in section 6.3.2.2 to generate A3G tetramer models, which can then be compared to our A3G scattering data. However, this work remains to be concluded. Nevertheless, from our SEC-SAXS experiments we can conclude that the ligand-free double-domain A3G exists in multiple states in solution that are yet to be structurally elucidated.

6.4.3 SAXS model of two-domain A3 in complex with ssDNA

6.4.3.1 SAXS model of a CCdZ30-40mer ssDNA

Several aspects contributed in the design and use of a 40mer dZ-containing oligonucleotide, to model the double-domain A3G-ssDNA complex using SAXS. As mentioned earlier in chapter 5, A3G has an intrinsic preference towards deamination of dC near the 3'-end of a CC<u>C</u>-motif (K_m ~ 570 µM for $A3G_{CTD}$ on 5'-ATTCC<u>C</u>AATT [94]). Incorporation of dZ in this position was reported to lead to significant inhibition of the $A3G_{CTD}$ activity (CCdZ) [136]. Contrary to $A3B_{CTD}$-AL1 (5FdZ-oligo ~ 2 µM, dZ ~ 8 µM, section 4.4.2), 5FdZ-containing ssDNA (5'-ATTCC<u>5FdZ</u>AATT, K_i ~ 71 ± 14 µM) did not show a marked improvement in the inhibition of A3G over dZ-containing ssDNA (5'-ATTCC<u>dZ</u>AATT, K_i ~ 53 ± 10 µM). Therefore, we decided to incorporate dZ into the CC<u>C</u>-motif to prevent deamination complications by the catalytically active A3G and to enhance the binding affinity of the ssDNA to the protein.

Moreover, it was reported that 20mer oligonucleotides had higher binding affinities to single domain A3 enzymes than shorter oligonucleotide sequences [135]. Therefore, a 40mer oligonucleotide was thought to better accommodate binding to two-domain A3 enzymes. Studies conducted by Dr Stefan Harjes demonstrated that extending the 10mer CCdZ- or CC5FdZ-oligos by a poly T_{30} tail at the 3'-end decreased the deamination rates of the active A3G in comparison to the 10mer oligonucleotides, consistent with an increased affinity for the longer oligonucleotide (illustrated in Figure 6.21). Placement of the CCdZ or CC5FdZ-motifs near the 5'-end of the 40mer ssDNA (CCdZT_{30}-oligo and CC5FdZT_{30}-oligo, Figure 6.21) led to significant inhibition of the full-length A3G activity, whereas incorporation of these motifs near the 3'-end (T_{30}CCdZ-oligo and T_{30}CC5FdZ-oligo,

Figure 6.21) did not greatly reduce the enzymatic activity. This implied that the specificity of these oligonucleotides to inhibit the two-domain A3G must be spatially oriented towards the CTD (5'-CCdZT$_{30}$ or 5'-CC5FdZT$_{30}$). These observation were also reported by Chelico et al [48]

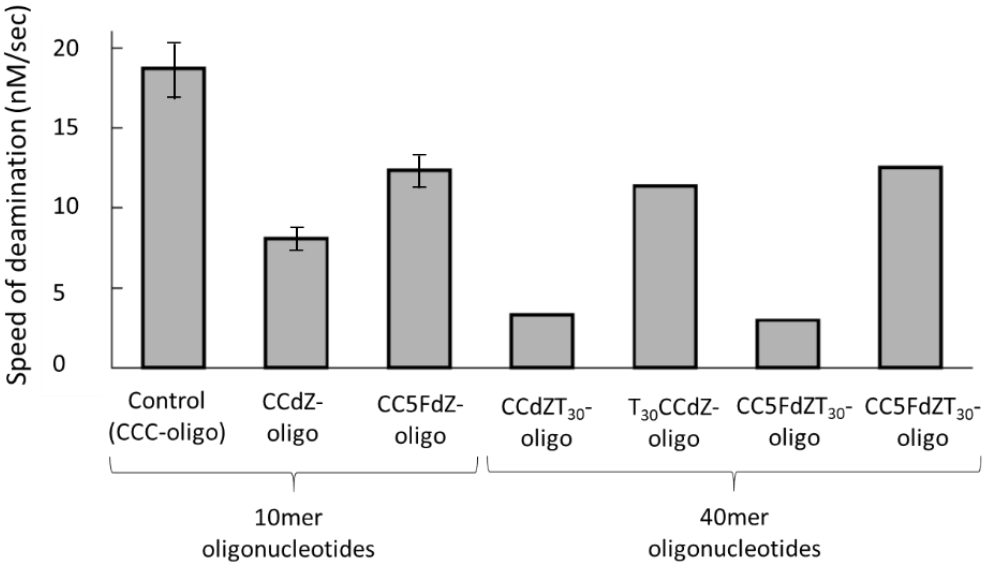


Figure 6.21 Qualitative screen of modified-oligos on the inhibition of A3Gfl-Wt catalysed deamination of the preferred dC substrate in 5'-ATTCCCAATT.

Plot of the initial speed of deamination of 500 µM 5'- ATTCCCAATT (target dC underlined) by the A3Gfl-Wt in the absence or presence of 50 µM CCdZ- or CC5FdZ-containing oligos. Experiments were performed using the ^{1}H-NMR-based inhibition assay (see section 2.5.3) in a pH 6.0 buffer (100 mM NaCl, 50 mM sodium phosphate buffer, pH 6.0, 10 % D_2O, 1 mM citrate, 50 µM DSS) at 298 K. Experiments with 10mer oligonucleotides were repeated multiple times and the mean values were plotted with error bars reported as SD, whereas single experiments were performed on 40mer oligonucleotides. Experiments were performed and results obtained by Dr. Stefan Harjes [136]

Furthermore, Harjes et al reported that the inhibition of the A3G$_{CTD}$ was comparable to the overall inhibition of the two-domain A3G [136], signifying that inhibition of the catalytic activity of the two-domain A3G was attributed to inhibition of the CTD rather than the NTD, which regardless lacks deamination activity [187]. Taken together, the specificity of modified-ssDNA to bind to double-domain A3 relies on the CTD and the ssDNA's orientation. Therefore, to study the structure of the double-domain A3G in complex with ssDNA, a 40mer CCdZT$_{30}$-oligo (5'-ATTCCdZAATTT$_{30}$, poly T$_{30}$ tail at the 3'-end) was selected to ensure binding to the CTD and inhibition of the catalytically active A3G.

Initially, the CCdZT$_{30}$-oligo (termed CCdZ-40mer) was examined using SEC-co-flow-SAXS to understand its dynamic structure in solution and to aid in modelling the A3G-ssDNA complex. The CCdZ-40mer oligo eluted off the SEC column as a single monodisperse species (Figure 6.24B), with a steady *R*g ~32 Å displayed across the peak in both the elution and SAXS profiles (see Figure 6.22A and Figure 6.24B). The double log plot resulted in a plateau at low *q* values, indicating that the sample was homogenous (see insert Figure 6.22B ii), which was further verified by a good fit to the linear regression in the Guinier plot (insert Figure 6.22B i).

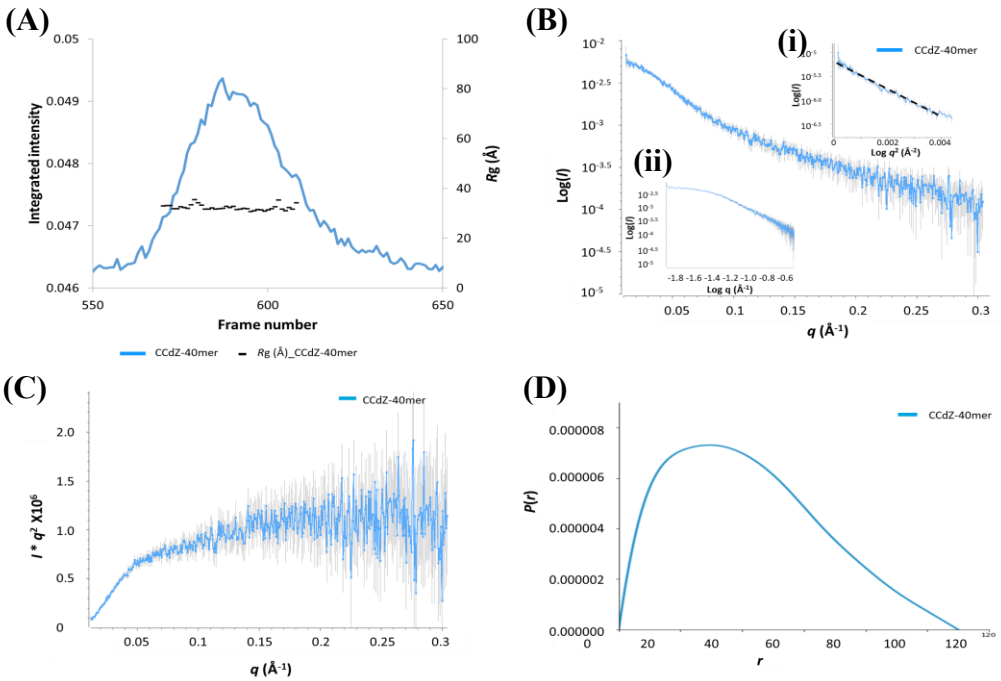


Figure 6.22 SEC-SAXS analysis of the CCdZ-40mer oligo.

SAXS analysis of the scattering data of the CCdZ-40mer oligo conducted using several programs from the PRIMUSQT ATSAS 2.8.3 suite [75]. (A) SAXS profile with *R*g trace values superimposed over profile, (B) scattering curve with inserts of the (i) Guinier plot and (ii) double log plot, (C) Kratky plot. (D), *P*(*r*) distribution plot. Experiments preformed using 0.68 mg/mL (~ 56 µM) of CCdZ-40mer ssDNA in A3G pH 6 kinetic buffer.

Estimation of the *R*g and *I*(0) values from the Guinier slope agreed well with values obtained by the independent *P*(*r*) method, further confirming the quality and relative size of the oligonucleotide listed in Table 6.9. From these parameters the estimated MW ~12-13 kDa of the oligonucleotide was found to be comparable to the expected MW ~12 kDa (Table 6.9). The Kratky profile indicated that the oligonucleotide adopted a highly flexible extended conformation, as values at $q > 0.1$ Å^{-1} did not return back to the baseline (Figure 6.22C). This was additionally supported by the *P*(*r*) plot, which was significantly

skewed to the right in comparison to the standard symmetrical bell-shaped histogram (Figure 6.22D).

Table 6.9 SAXS structural parameters of the CCdZ-40mer oligo sample

Structural parameters	CCdZ-40mer oligo
Guinier analysis	
$I(0)$ (cm^{-1})	0.00530 ± 0.00012
Rg (Å)	32.0 ± 1.0
$q$min ($Å^{-1}$)	0.0165
$q \cdot R$g max	1.3
Coefficient of correlation (R^2)	0.97
P(r) analysis	
$I(0)$ (cm^{-1})	0.005260 ± 0.000008
Rg (Å)	32.2 ± 0.3
D_{max} (Å)	118.0
q range ($Å^{-1}$)	0.0165 - 0.2975
Quality estimate	0.78
Porod volume ($Å^3$)	20,700
MW (Porod Volume*0.6) (Da) (ratio to expected subunit 12,084 Da)	12,420 (1.03)
MW (from $I(0)$ [178]) (Da) (ratio to expected subunit 12,084 Da)	13,150 (1.09)

Reconstruction of the scattering profile using 3D envelope modelling (NSD ~ 0.875) demonstrated, remarkably, that we can visualise the grooves within the oligonucleotide. The envelope model mimics a dumbbell, which is consistent with the shape of the $P(r)$ curve (observed in Figure 6.22D), as described in literature [209]. This model illustrates that the oligonucleotide is made up of four helical turns, consistent with the number of base pairs in the 40mer ssDNA (ten base pairs is equivalent to one helical turn). Interestingly, we observe that the CCdZ-40mer oligo forms a kink in the middle of the sequence (Figure 6.23B i). This may be due to the adenosine near the modification, which may influence sugar puckering leading to stacking of the nucleotides above and below.

Based on the envelope model, the CCdZ-40mer oligo adopts the standard B-form DNA structure, even though it is single stranded, under our conditions in solution, which is a less compact helical structure in comparison to the standard compact A-form (Figure 6.23B). Therefore, the averaged envelope model was superimposed with a B-form 40mer ssDNA structure (designed using the *make na* server http://structure.usc.edu/make-na/server.html), and modelled with a kink, which gave a nice visual fit (Figure 6.23B ii). To validate this model, rigid body modelling was performed by comparing back-calculated 1D scattering profiles of the designed B-form DNA structure to the experimentally derived CCdZ-40mer scattering data using both FoXS [74, 226] and CRYSOL [75]. The scattering data showed a good visual fit to the B-DNA model, especially in the *q* regions between 0.02 – 0.2 Å^{-1} as illustrated in Figure 6.23 and summarised in Table 6.10. Therefore, the in-solution SAXS-based model of the CCdZ-40mer oligo is comparable to the rigid B-DNA model structure.

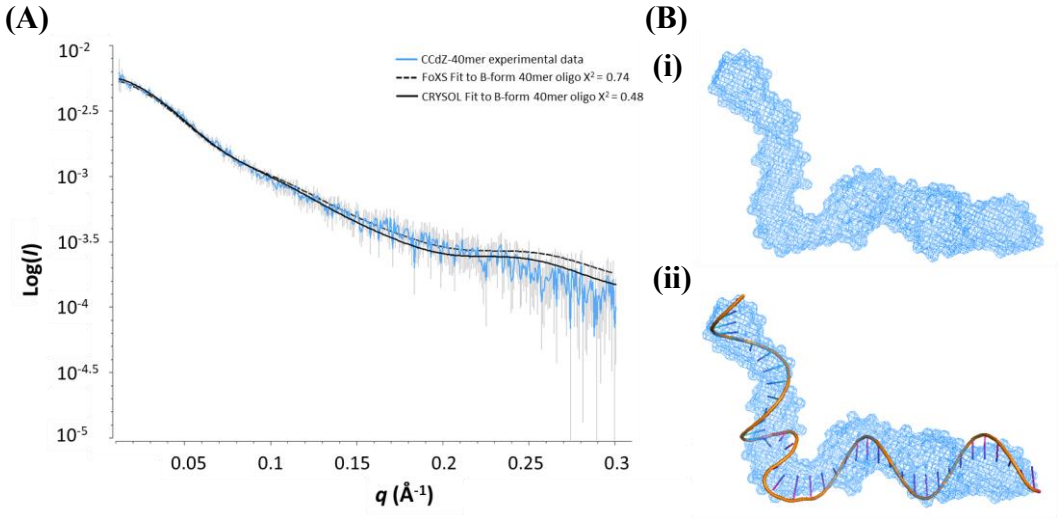


Figure 6.23 Fitting CCdZ-40mer ssDNA experimental data.

(A) Rigid body modelling of CCdZ-40mer oligo data to a B-form 40mer oligo model structure using FoXS and CRYSOL (B) (i) Averaged envelope model of CCdZ-40mer oligo using *P*1 symmetry using DAMMIF and refined with DAMAVER and DAMFILT (ATSAS 2.8.3 suite) [75]. (ii) Envelope model superimposed with B-form DNA structure.

Table 6.10 SAXS fitting and modelling parameters of the CCdZ-40mer ssDNA sample

Modelling parameters	CCdZ-40mer oligo
Ab intio restoration	
DAMMIF (default parameters, 10 calculations)	
q range ($Å^{-1}$) for fitting	0.0165 - 0.2975
Symmetry	*P*1
NSD (standard deviation)	0.875 (0.052)
Resolution (from SASRES) (Å)	36 ± 3
Structure modelling	
40mer ssDNA B-form model	
FoXS, Chi^2	0.74
CRYSOL, Chi^2	0.43

6.4.3.2 SAXS model of the active two-domain A3G in the presence of dZ-containing ssDNA

To elucidate a model for the two-domain A3 in complex with ssDNA, SEC-SAXS experiments were performed on A3Gfl-Wt in the presence of CCdZ-40mer oligo (described earlier 6.4.3.1) in A3G kinetic pH 6.0 buffer. However, we observed that the sample was not homogeneous and the eluents were not well resolved, similar to the ligand-free A3Gfl-Wt sample (see section 6.4.2). Focusing on the elution profiles, the *R*g trace was observed to be variable across the protein elution peaks, denoting that multiple A3G species were present in the sample (*R*g ~ 30 - 54 Å across 460 - 590 sec, Figure 6.24C), as detailed earlier in section 6.4.2. However, *R*g remains constant over the region where the oligo elutes (*R*g ~ 33 Å, from 597 – 645 seconds, Figure 6.24C). However, when we compared the absorbance maximum of the eluted ssDNA peak from this sample (A_{280} ~ 1.02, Figure 6.24C) to that of ssDNA control sample (A_{280} ~ 1.53, Figure 6.24B), we found that the absorbance significantly decreased by ~ 30 %. This decrease was reciprocated by an increase in the overall absorbance of the eluted protein fractions, establishing that a protein-DNA complex had formed. However, as the A3Gfl-Wt protein existed in multiple conformation under our conditions (detailed earlier in section 6.4.2), it was unclear from the elution profile which A3G form was in complex with the oligonucleotide (further discussed later in Figure 6.27). Therefore, Gaussian decomposition analysis was performed on the region corresponding to the protein elution

(frame numbers 385 - 650, Figure 6.25A), to extract the scattering curves for each A3G component from the A3Gfl-Wt/CCdZ-40mer sample.

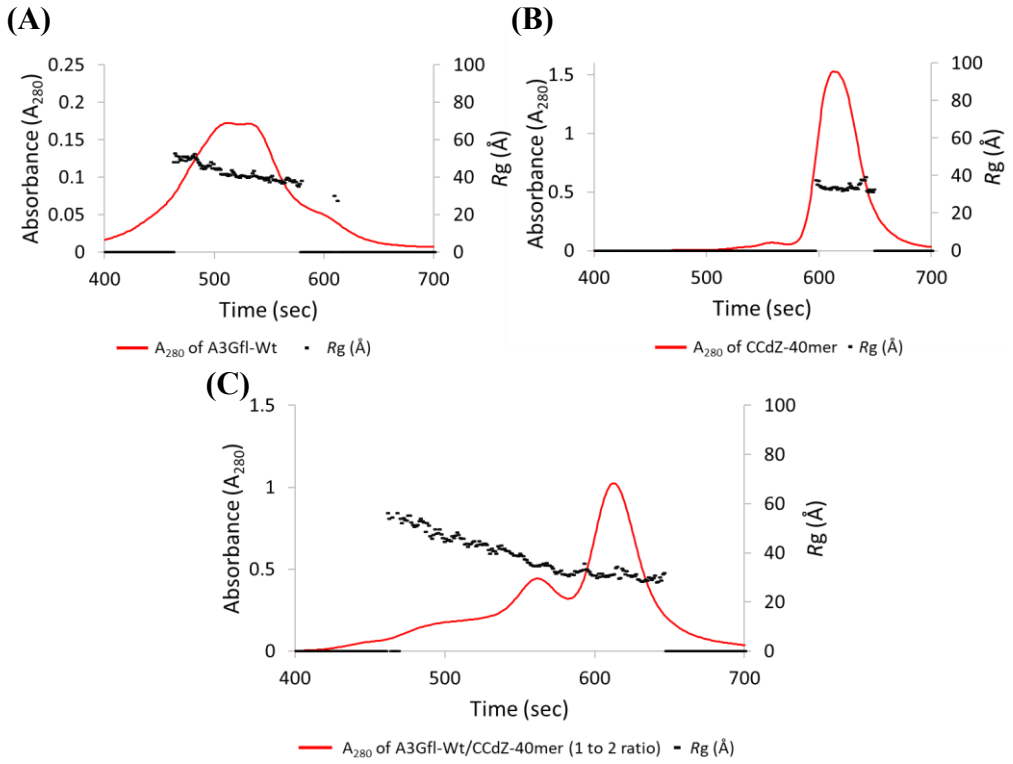


Figure 6.24 SEC elution profiles of A3Gfl-Wt with and without ssDNA.

SEC-FPLC performed using (A) A3Gfl-Wt at 2.5 mg/mL (~ 56 µM), (B) 0.68 mg/mL (~ 56 µM) of CCdZ-40mer oligo and, (C) A3Gfl-Wt with CCdZ-40mer ssDNA at a 1 to 2 ratio. Samples were run at 25 °C through a SEC column at a flow rate of 0.2 mL/min in A3G pH 6 kinetic buffer (50 mM phosphate pH 6.0, 200 mM NaCl, 2 mM β-ME, 5% glycerol, 200 µM Na_2-EDTA).

Initially, SVD and EFA analyses were conducted using SVD/EFA BioXTAS RAW [105] to establish the number and boundaries of each scattering species, as described earlier in section 6.1.1. SVD estimated that there were four components in this region (refer to Figure 8.13). EFA analysis integrated the boundaries of each species (EFA 1: 385-526, EFA 2: 446-559, EFA 3: 485-583, EFA 4: 477-590, Figure 6.25B). Plotting these as individual peaks we found that the components in EFA 2, 3, and 4 were major contributors to the scattering pattern, while EFA 1 was barely present (Figure 6.25B). To further isolate the 1D-scattering curves of each component Gaussian decomposition analysis (US-SOMO) were performed by Dr. Timothy Ryan (SAXS/WAXS beamline scientist, Australian Synchrotron) as described earlier in section 6.4.2. Four Gaussian functions were fitted to the data (Figure 6.25C), then converted into four 1D-scattering curves of each component (termed species 1-4).

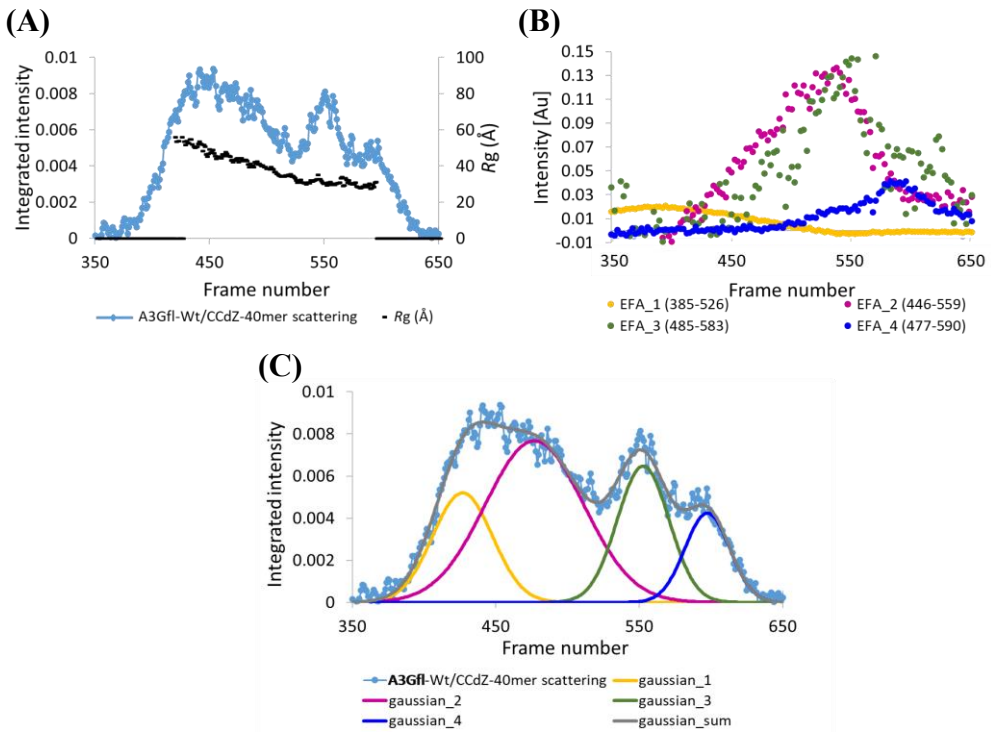


Figure 6.25 Deconvolution of SAXS data of the A3Gfl-Wt/CCdZ 40mer sample.

SEC-SAXS analysis of A3Gfl-Wt/CCdZ-40mer sample (A) SAXS profile with *R*g trace values superimposed over profile. (B) EFA analysis using BioXTAS RAW (C) Gaussian decomposition analysis using US-SOMO. Experiment conducted using 2.5 mg/mL A3Gfl-Wt in a 1 to 2 ratio with CCdZ-40mer oligo in A3G pH 6.0 kinetic buffer at 25 °C.

Initially, we established using the double log plot that all the derived scattering curves contained only one scattering species (Figure 6.26B), as explained in section 6.1.1. The Kratky plots of species 1-3 presented characteristic bell-shaped peaks at low *q* implying that the proteins were globular, though the structures also had flexible regions as indicated by high *q* data not completely returning to baseline (Figure 6.26C). These results were again consistent with the skewed shape of the *P*(*r*) curve from a standard bell-shaped curve, implying these components had elongated shapes (Figure 6.26D). Interestingly, the Kratky plot of species 4 had a broader peak that downturned at low *q*, which is typical of a multidomain structure with a flexible linker (Figure 6.26C). However, due to the low abundance of species 1 as indicated by its low signal intensity (Figure 6.25B) no accurate parameters could be derived for this species (see Figure 6.26B insert). Nevertheless, the Kratky plot of species 1 indicated that it was twice as big as species 2, based on the peak maximum positions as described earlier in section 6.4.2. Therefore, we focused on deriving structural parameters for species 2-4, as listed in Table 6.11.

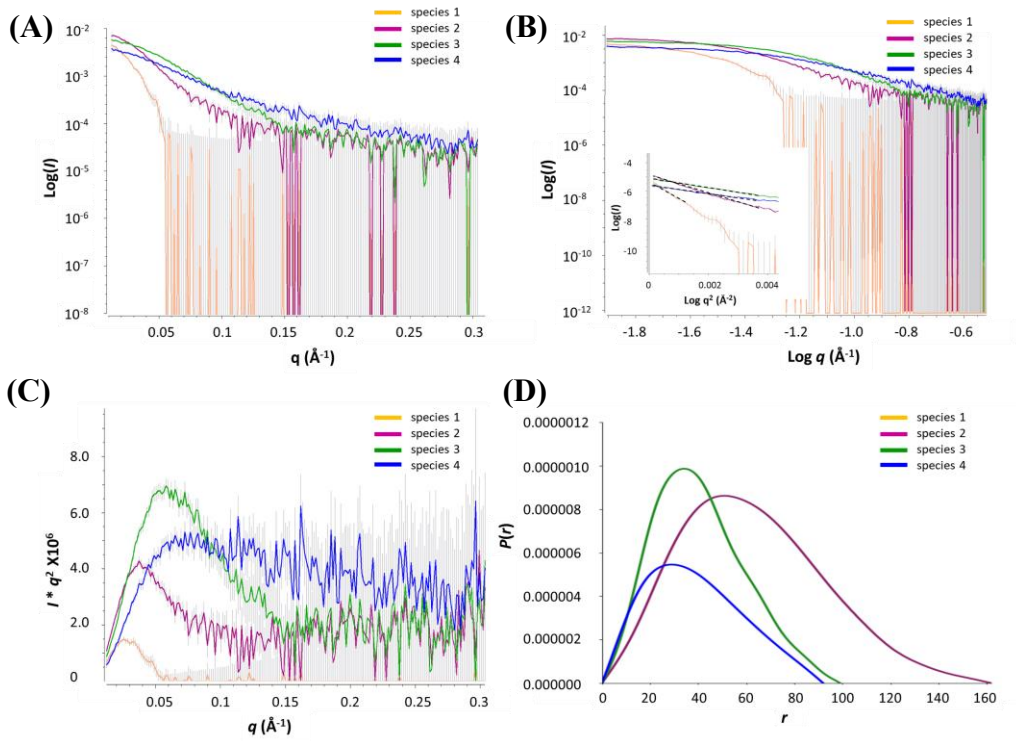


Figure 6.26 SEC-SAXS analysis of A3Gfl-W/CCdZ-40mer deconvoluted species.

Analysis of the scattering data of the A3Gfl-Wt/CCdZ-40mer deconvoluted species was conducted using several programs from the PRIMUSQT ATSAS 2.8.3 suite [75]. (A) 1D-scattering curves of each species (B), double log plot with insert of a Guinier plot, (C) Kratky plot. (D) *P*(*r*) distribution plot. Experiments were conducted using 2.5 mg/mL of A3Gfl-Wt in A3G pH 6.0 kinetic buffer.

Table 6.11 SAXS structural parameters of the A3Gfl_Wt in presence of CCdZ-40mer oligo

Structural parameters	Species 2	Species 3	Species 4
Guinier analysis			
I(0) (cm^{-1})	0.00812 ± 0.00018	0.00492 ± 0.00093	0.00261 ± 0.00092
*R*g (Å)	47.0 ± 1.4	33.6 ± 1.2	29.8 ± 1.7
*q*min ($Å^{-1}$)	0.0139	0.0183	0.0183
q·*R*g max	1.3	1.3	1.3
Coefficient of correlation (R^2)	0.98	0.45	0.57
P(r) analysis			
I(0) (cm^{-1})	0.00834 ± 0.00012	0.00496 ± 0.00069	0.00244 ± 0.00044
*R*g (Å)	48.4 ± 0.6	32.7 ± 0.5	30.4 ± 0.4
D_{max} (Å)	162.1	100.1	92.9
q range ($Å^{-1}$)	0.0139 - 0.3052	0.0183 - 0.3052	0.0183 - 0.3052
Quality estimate	0.89	0.80	0.79
Porod volume ($Å^3$)	379,000	117,000	62,700
MW (Porod Volume*0.6) (Da) (ratio to expected subunit 46,408 Da)	227,400 (4.9)	70,200 (1.5)	37,620 (0.8)

6.4.3.2.1 Analysis of species 2 from A3Gfl-Wt/CCdZ-40mer SEC-SAXS

Based on the estimated MW, we deduced that species 2 corresponded to an A3G tetramer made up of four two-domain subunits (Table 6.11). We noted, that the size (MW, *R*g, D_{max}, Table 6.11) of this species was slightly larger (estimate of 4.9 subunits) than the ligand-free A3G-tetramer species (estimate of 4.4 subunits, Table 6.7), indicating that this component was in complex with an oligonucleotide. To verify this notation, the A_{260}/A_{280} ratio of the A3Gfl-Wt sample was compared to the A3Gfl-Wt/CCdZ-40mer sample, as described earlier in section 6.3.4. Interestingly, the A_{260}/A_{280} ratio of the A3Gfl-Wt/CCdZ-40mer sample appeared to be larger than the ratio of the ligand-free protein across the entire elution profile as displayed in Figure 6.27, illustrating that the our A3G species 2 (A3G tetramer), and all other species (1, 3, and 4) elute along with DNA. This signified that each of these species forms a complex with the CCdZ-40mer oligo.

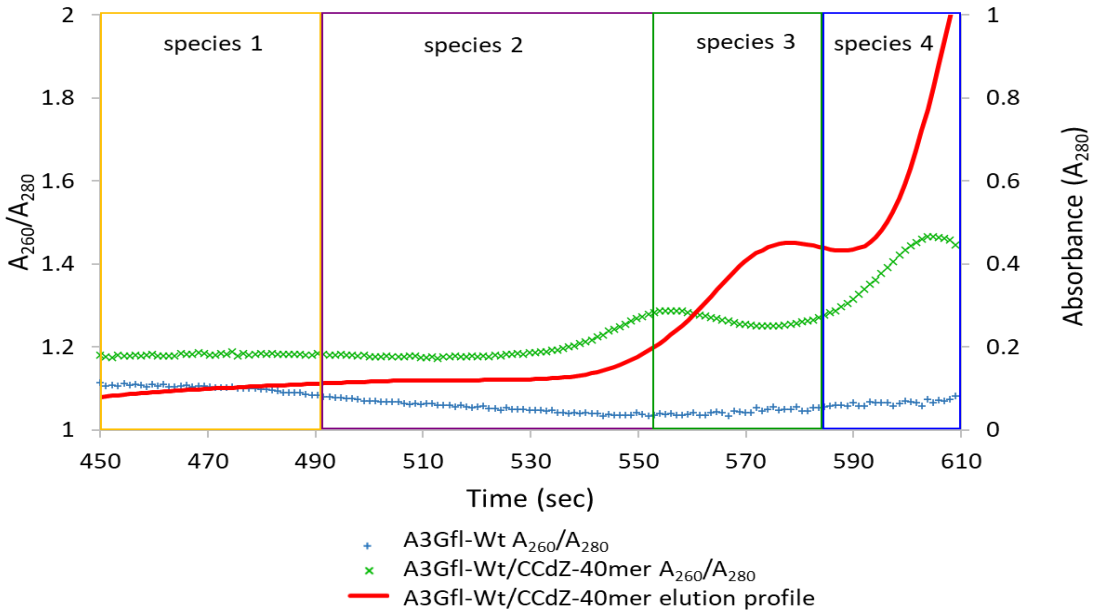


Figure 6.27 Ratio of A_{260}/A_{280} to assess the presence of DNA during the elution of the A3Gfl-Wt/CCdZ-40mer sample.

A_{260}/A_{280} ratio of the A3Gfl-Wt (blue) and A3Gfl-Wt/CCdZ-40mer (green) samples. Overlaid with the A_{280} elution profile of the A3Gfl-Wt/CCdZ-40mer sample (red) to display the boundaries of each species (species 1, 2, 3, and 4).

Based on differences between the MW of species 2 (MW ~ 227 kDa, Table 6.11) and the ligand-free A3G tetramer (MW ~ 203 kDa, Table 6.7), we can deduce that two oligonucleotides (24 kDa ~ two 12 kDa CCdZ-40mer oligos, Table 6.9) form a complex with the A3G tetrameric species 2. To help elucidate the 3D structure of this species, *ab initio* shape reconstruction (DAMMIF) was performed using *P*4 symmetry as this sample was determined to be tetrameric. A mean NSD score of 0.578 for the averaged envelope model was derived, which indicated very good consistency for the ensemble (acceptable NSD ≤ 0.8) (Table 8.12). The envelope model was globular and resembled a cross as illustrated in Figure 6.28A. Superimposing this model with the envelope model of the previously identified ligand-free A3G-tetramer, we find that the ligand-free envelope has a rather nice fit to this model. However, two noticeable areas either side of the docked ligand-free A3G tetramer model appeared to be unoccupied as visualised in Figure 6.28B. These empty regions indicate the binding of two oligonucleotides, confirming earlier interpretations, whereby the A3G tetramer (species 2) complexes with two CCdZ-40mer oligonucleotides in solution, under our conditions.

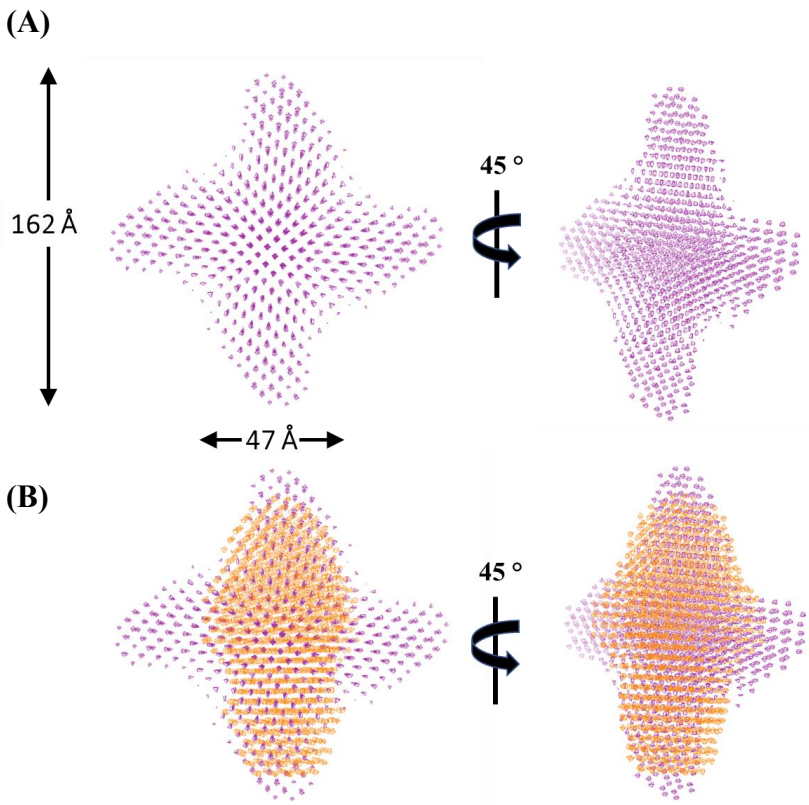


Figure 6.28 Envelope models of the A3G tetramer with and without CCdZ-40mer oligo.
(A) Envelope model of the A3G tetramer/CCdZ-40mer oligo (purple) generated under P4 symmetry (DAMMIF, ATSAS 2.8.3 suite) [12], with D_{max} ~ 162 Å and *R*g ~ 47 Å. (B) A3G tetramer/CCdZ-40mer envelope model (orange) superimposed with ligand-free A3G tetramer.

Interestingly, upon binding of the oligonucleotides to the A3G tetramer we see that the protein seems to elongate slightly in comparison to the ligand-free form (Figure 6.28B), which was consistent with a 17 Å increase in D_{max} (Table 6.11). The action of binding these ssDNAs to the tetrameric protein may cause structural rearrangements in the A3G tetramer leading to the observed elongation of the molecule. However, as we do not have high-resolution A3G tetrameric structures we cannot conclude this theory.

6.4.3.2.2 Analysis of species 3 and 4 from A3Gfl-Wt/CCdZ-40mer SEC-SAXS

Estimation of the MWs of species 3 and 4 from the Porod volumes seemed to be underestimated and inaccurate (Table 6.11), due to a low signal-to-noise ratio following deconvolution of the scattering data. Furthermore, calculating MWs using the $I(0)$ method [178] was difficult as estimation of protein concentration estimates of each species presented large inaccuracies (peak overlaps). Nevertheless, based on the peak maxima position of the Kratky plots, we can estimate the oligomeric state of these species. By comparison, our DNA bound A3G tetramer species 2 (q ~0.036 Å^{-1}) was determined to be ~ 1.7 times larger than species 3 (q ~0.061 Å^{-1}), and ~ 2.6 times larger than species 4 (q ~0.092 Å^{-1}) (Figure 6.26C).

As the DNA bound A3G tetramer (species 2) was estimated to be made up of 4.9 subunits (four A3G two-domain subunits and two oligos) (Table 6.11), using this information we estimated the number of subunits in species 3 and 4 (number of subunits in species 2/estimated difference of size). Species 3 was calculated to be composed of approximately 2.9 subunits (4.9/1.7), whereas species 4 was estimated to be comprised of ~ 1.9 subunits. As shown earlier in Figure 6.27, both species 3 and 4 were indicated to complex with our CCdZ-40mer oligo. Therefore, based on these annotations we can conclude that species 3 corresponds to an A3G dimer in complex with CCdZ-40mer, whereas species 4 may in fact be an A3G monomer (single double-domain) in complex with DNA. However, whether one or more oligonucleotides is bound to these species is difficult to conclude from these data.

To endorse these claims, rigid body modelling was performed by comparing the experimentally derived scattering data of species 3 and 4 to back-calculated 1D scattering profiles of the reported A3G dimer atomic structures (PDB; 6p40, 6p3z, 6p3x, and 6p3y) [286] using FoXS [74, 226] and CRYSOL [75]. As anticipated, species 3 had an overall better fit to the dimeric A3G atomic structures ($Chi^2 \leq 1.28$, Figure 6.29A) than species 4 (high

Chi^2, Figure 6.29B) as listed in Table 8.13, further supporting our model that species 3 was an in fact an A3G dimer. We observed that species 3 had a slightly better visual fit against the A3G dimeric structure 6p40, especially at low *q* regions (0 -0.1 Å^{-1}, Figure 6.29A), in contrast to the other A3G dimer structures. However, this was not clearly indicated by its Chi^2 scores < 1, suggesting that Chi^2 values of the other A3G dimer structures (PDB, 6p3z, 6p3x, and 6p3y) result in better fits (see Table 8.13). Moreover, at $q > 0.12$ Å^{-1} the scattering data of species 3 did not fit nicely to the structural models (Figure 6.29A), consistent with the additional scattering of our CCdZ-40mer oligo, which was bound to our A3G tetramer (species 3).

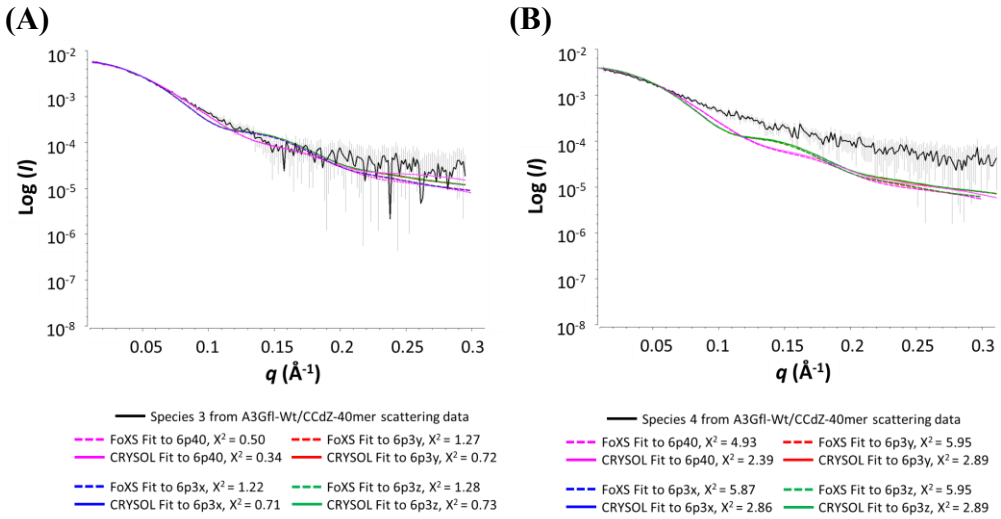


Figure 6.29 Rigid body modelling of A3Gfl-Wt/CCdZ-40mer species 3 (A3G dimer) and species 4 (A3G monomer) scattering data to known A3Gfl dimer crystal structures.

Rigid body modelling of the A3G dimer/CCdZ-40mer (species 3) and (B) A3G monomer/CCdZ-40mer (species 4) to A3Gfl dimer PDB structures (6p40, 6p3x, 6p3y, and 6p3z [286]) using FoXS and CRYSOL.

On this basis, envelope models were generated using a *P*2 symmetry for species 3, and *P*1 symmetry for species 4 (DAMMIF) as described in section 6.1.1, to visualise the low-resolution model of each species. Using this method, the averaged envelope model for species 3 resulted in a mean NSD score of 0.937 (acceptable NSD ≤ 0.8), indicating that the model was not an accurate representation of this species (see model in Figure 8.14A). Therefore, no further modelling was continued for this component. However, a very good mean NSD score of 0.559 was obtained for the envelope model of species 4, which indicated consistency of this ensemble. Docking half of the atomic A3G dimer structure as an A3G full-length monomer (A3G two-domain structures all similar so used PDB,

6p3z) into the envelope model of species 4, we observed that this structure fits only into part of the envelope model leaving a large region of electron density unoccupied on one side of the envelope (Figure 6.30A insert). This was consistent with a poor fit of the species 4 scattering data to the A3G-full length structure, as illustrated by the high Chi^2 scores and fit in Figure 6.30A. The unoccupied region indicated the binding location of the CCdZ-40mer ssDNA, and suggested that a single oligonucleotide binds to the two-domain A3G monomer (Figure 6.30A insert).

Therefore, we manually superimposed our 40mer B-form ssDNA (displayed in Figure 8.15C) into this vacant space and positioned it into the active-site in several ways. Using this method, we derived several models that were then assessed against the scattering data of species 4 using rigid body modelling (see Figure 8.15). We first noted that the presence of our oligonucleotide within the structure significantly improved the overall fit of species 4 scattering data (Figure 6.30B), in contrast to the ligand-free structure (Figure 6.30A). Overall, our generated atomic models had an okay fit to the scattering data (Figure 8.14), noting that model 4 resulted in a better overall fit (as indicated by the Chi^2 scores, Figure 8.14D).

However, in all our models (see Figure 8.15) the oligonucleotide appeared to be largely bound to only one of the domains with less contact to the other domain. From these models, we propose that the oligonucleotide was bound to the CTD and was not specifically associating with the NTD. From these observations we can envisage that our model may resemble the in-solution structure of an active two-domain A3G in complex with a ssDNA inhibitor. As described earlier in section 6.4.3.1, inhibition of the two-domain A3G was reported to be attributed to the specificity of the dZ/5FdZ-containing oligos to bind to the catalytic CTD rather than the NTD. Based on these data reported by Dr. Stefan Harjes (see Figure 6.21), it is reasonable to conclude that the fit of our SAXS model supports inhibition. whereby dZ would be bound in the active site of the CTD (Figure 6.30B), in a similar manner as structures of the previously reported inactive single-domain A3 in complex with ssDNA containing substrate dC [93, 239].

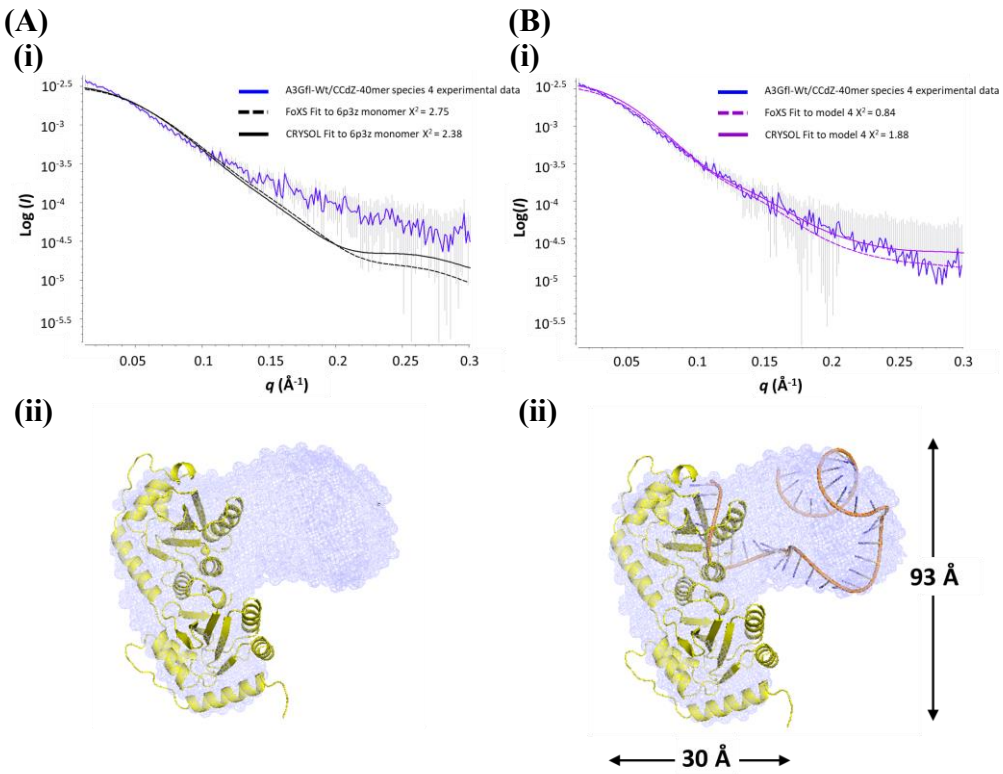


Figure 6.30 Fitting SAXS profile of A3G/CCdZ-40mer species 4 to model structure 4.

(A) Fit of 6p3z monomer to SAXS data for species 4 (B) Fit of 6pz3 monomer plus model for CCdZ-40mer to SAXS data for species 4. Inserts show the models docked in the molecular envelopes calculated from the SAXS data .

Interestingly, our model showed that the long poly-T tail at the 3'-end of the oligonucleotide remained fairly flexible. Several studies indicated that the presence of the NTD enhances the deamination efficacy in two-domain A3 over the CTD alone, suggesting its importance in binding ssDNA [24, 78, 88, 91, 187, 189, 247]. Based on these observations, we expected to observe that the 3'-end long tail of the oligonucleotide would be associated with the NTD, supporting the ~2-fold increase in the inhibition of the enzyme by the long CCdZT_{30}-oligo versus the short 10mer CCdZ-oligo (Figure 6.21, noting that 40mer oligonucleotides were assessed with only a single experiment). However, from our SAXS data, we found that the long tail had limited interactions with the NTD and it looped away from the protein (Figure 6.30B). This suggested that in-solution binding of the ssDNA to the NTD may be dynamic and loosely fixed, while binding of the oligonucleotide into the CTD is more specific. To validate these observations further refinement with molecular dynamics and elucidation with higher resolution structures would be advantageous to understand these mechanisms. However,

the structural parameters and models obtained from SEC-SAXS provide a platform for further structural studies using chemically modified oligonucleotide inhibitors.

6.5 Conclusions

Here, we report the first SAXS models of catalytically active A3s in complex with dZ-containing ssDNA inhibitors. Initially, we studied the catalytically active $A3B_{CTD}$-AL1 in the presence of a dZ-containing oligo. Our model indicated that dZ-containing ssDNA was bound to the $A3B_{CTD}$-AL1, while the 3'-end of the oligonucleotide remained rather flexible (Figure 6.6). This low-resolution model was found to be comparable to the previously reported inactive substrate-bound $A3B_{CTD}$ crystal structure (PDB, 5TD5 [239]) (Figure 6.6). This result further suggests that our dZ-containing inhibitor binds into the active-site of $A3B_{CTD}$ in a similar manner to the dC substrate oligonucleotide, as supported by our inhibition mechanism (described in section 3.4.2).

Under our conditions, the ligand-free $A3B_{CTD}$-AL1 was detected to elute from the SEC column as two species, the monomeric $A3B_{CTD}$-AL1 form and a secondary larger multimer. Using SEC-SAXS, we were able to elucidate that this secondary eluting species was in fact an $A3B_{CTD}$-AL1 homodimer in solution. A model structure that was generated was in support of a dimer interface that was formed through several weak interactions, but stability was largely attributed to a covalent disulfide bond. Interestingly, this interface seemed to mimic a previously reported A3A dimer (PDB, 4xxo [23]), which was not surprising as our $A3B_{CTD}$-AL1 variant contained a loop from A3A (A3A loop 1) that was involved in the formation of this interface. However, examination of our $A3B_{CTD}$-AL1 dimer in the presence of the dZ-oligo inhibitor, revealed that contrary to its observed catalytic activity (Figure 8.2) the dimer was in fact not bound to the oligonucleotide (Figure 6.16), seemingly at odds with observed catalytic activity of this species, described earlier [22-23, 27, 79, 118, 233]. However, we always observed a small amount of monomeric $A3B_{CTD}$-AL1 with the dimeric species. The observed catalytic activity is ascribed to this monomer species, noting the presence of the redox-active citrate in the citrate-phosphate buffer used.

Under our conditions, ligand-free two-domain A3G eluted from the SEC column in multiple oligomeric states but was largely found to exist as higher molecular weight (HMW) species, consistent with published studies [48, 50, 55]. Lower molecular weight (LMW) species were present but in very low quantities, which could not be quantified

(Figure 6.18A and D). These various multimeric states may aid with the stability of two-domain A3 proteins in regulating their catalytic activity [23, 48, 50, 55, 150, 276]. In particular, using derived SAXS parameters we identified that an A3G tetramer was the most prominent eluting species. This species formed from four two-domain A3G molecules oligomerising together in a compact formation (MW ~ 200 kDa), potentially through surface associations like our reported $A3B_{CTD}$-AL1 dimer (discussed in section 6.3.2.2), though this remains unclear.

In the presence of the CCdZ-40mer ssDNA, several LMW A3G conformations (a dimer and monomer species) became much more prominent, while the higher HMW formations decreased slightly (see Figure 6.18 and Figure 6.25). This suggested that the higher order multimeric species may be converted into these smaller configurations due to interactions with the ssDNA, as all our eluted A3G species were identified to be in complex with ssDNA, as indicated upon examination of the A_{260}/A_{280} ratio. From this we propose that the binding of the negatively charged ssDNA to the larger A3G conformations may displace the interface interactions between the A3G molecules.

The previously identified ligand-free A3G tetramer was determined to be capable of forming a complex with two CCdZ-40mer oligonucleotides. However, a model structure was difficult to elucidate as the formation of the binding interfaces between the A3G molecules was unknown. Two LMW A3G conformations were also detected. An A3G dimer was identified as its scattering data had a passable fit to the reported A3G dimer crystal structures [286], although this dimer formed a complex with ssDNA. However, due to its low concentration a SAXS model could not be elucidated. An A3G monomer complexed with a CCdZ-40mer oligonucleotide was discovered, revealing the first full-length catalytically active A3 model structure in complex with a ssDNA inhibitor. Our model showed that the modified-oligonucleotide was largely anchored into the CTD, consistent with the observed inhibition of the enzyme's activity and higher binding affinity for dZ (Figure 6.21). In solution the long poly-T tail seemed to be flexible and had a limited interaction with the NTD. Interestingly, this observation deviated from earlier studies [24, 78, 88, 91, 187, 189, 247], which implied that the ssDNA was bound to NTD enhancing the deamination efficiency of full-length A3, in comparison to CTD alone. However, based on the A3G inhibition assays reported by Harjes, it is notable to mention that the inhibition of the full-length A3G enzyme by our dZ/5FdZ-containing oligonucleotides was determined to be largely due to the inhibition of the CTD [136]. This

seemed consistent with our derived A3Gfl-Wt/CCdZ-40mer bound SAXS model, where the modified oligonucleotide has a higher affinity towards the CTD active site. Based on these observations, we propose that the ssDNA may initially bind to the domains of A3 in an allosteric manner. DNA binding to the NTD causes a conformational change in the CTD which helps position the dZ close to the CTD active-site. Upon, specific binding of the dZ into the CTD active site, the dZ is converted into a transition state of deamination where it remains bound. The nucleotides flanking the target dZ stack on top of one another to stabilise the overall conformation, as observed in the single domain A3-ssDNA complex structures [239]. These SAXS structural parameters and models may help enlighten further structural studies using modified-oligonucleotides.

7. Final conclusions and future directions

7.1 Summary of findings

Cancer recurrence, metastases, and drug resistance are some of the consequences of the mutagenic action of the A3B. Therefore, the selective inhibition of A3B may prevent rapid cancer evolution, while retaining the other A3 family members involved in the innate immunity. Until now, no selective A3B inhibitors have been described. In this research, we present a novel approach to inhibit the A3B's catalytic activity using chemically modified ssDNA containing cytosine analogue nucleosides that mimic the transition state of deamination. Here, we report the first selective ssDNA-based A3B inhibitor. Furthermore, we illustrate the first SAXS models of active $A3B_{CTD}$ and two-domain A3 in complex with our modified-ssDNA inhibitor, which will help further structural studies.

Coherent with our rationale, incorporating cytosine analogue nucleosides in place of the target dC in the A3s' preferred ssDNA sequences led to the inhibition of the catalytic activity of our $A3B_{CTD}$ enzyme, as well as A3A, $A3G_{CTD}$, and a full-length A3G as reported us and our collaborators [136]. These are the first substrate-like ssDNA A3 competitive inhibitors that utilise the preference of A3s to bind and deaminate cytosines in ssDNA [15, 26, 48]. The CDA prefers to target free dC or cytosine analogues over ones incorporated in ssDNA [289], so our ssDNA-based inhibitors would not interfere with CDA's functions. By embedding dZ and 5FdZ into the A3B's preferred deamination sequence motif (5'-TC) in a short ssDNA, we illustrated that the $A3B_{CTD}$ catalytic activity was inhibited with low micromolar inhibition constants (see Table 4.1). In particular, the dZ-containing ssDNA had an ~30-fold inhibition potency against $A3B_{CTD}$, whereas the 5FdZ-containing oligonucleotides had a higher potency close to 100-fold. This trend was also observed for the closely related A3A [136] and CDA against zebularine and 5-fluorozebularine [6]. These results indicate that the structure of the cytosine analogues dZ and 5FdZ determine the inhibitory potential against the A3 enzymes. Notably, the N3 of cytosine analogue was required to be protonated by the conserved catalytic glutamate in the A3s' active site to facilitate the inhibitory mechanism. This protonation authorises the nucleophilic attack of C4 of the cytosine analogue by an activated Zn^{2+}-coordinated water molecule that is necessary for the formation of a tetrahedral transition-state analogue of cytosine deamination (Figure 3.10). This is consistent with zebularine inhibition CDA [77]. As indicated when the $A3B_{CTD}$ catalytic glutamate was mutated to alanine (E255A) it caused significant reduction in ssDNA binding (Figure 3.11).

Most A3s (except A3G) have the same intrinsic preference of deaminating a cytosine preceded by thymidine, our modified substrate-like inhibitors based on a 5'-TC motif would not be selective towards the actions of A3B alone, as noted in the inhibition of the A3A [136]. The A3A and A3B were reported to deaminate different target dC within the suboptimal CCC-motif [37]. A3A (and A3G) prefers to target the dC at the 3'-end followed by a purine (dA or dG) [37, 41, 122-123, 207, 252] (5'-CCC, where underlined is the preferred dC), whereas A3B prefers the dC at the 5'-end flanked by a thymidine [36-37] (5'-CCC-motif, where underlined is the preferred dC). We exploited this differential target pattern, to obtain a selective A3B inhibitor. We noted that nucleotides adjacent to our cytosine analogue (dZ) in the CCC-motif influenced the specificity of the inhibitor. Utilising these observations, placement of dZ in place of the A3B's preferred dC in this motif (5'-dZCC) led to the first selective substrate-like inhibitor of A3B with a low micromolar inhibition constant (Table 5.4). This selective inhibitor had a similar inhibition efficiency of ~30-fold as our preferred non-selective dZ-containing inhibitor (in the 5'-TC motif), further supporting that the structure of the cytosine analogue defines the inhibitory potential. However, to date the cellular effects of our inhibitors are unknown and would need to be evaluated.

When this research began in 2015 there were no reported structures of the A3B; therefore, we attempted to determine its structure to better enlighten drug design. Using solution-state NMR experiments we aimed to elucidate the structure of a single domain $A3B_{CTD}$. However, during our study the ligand-free $A3B_{CTD}$ structures [37, 240] and an inactive $A3B_{CTD}$-ssDNA structure [37, 239] were solved. Nevertheless, we were able to assign the resonances of a ligand-free weakly active $A3B_{CTD}$-QM-ΔL3 and made use of the assignment by establishing the experimental conditions of chemical shift perturbation assays. Although the substrate-bound $A3B_{CTD}$ structure has been established [37, 239], we can use our chemical shift perturbation experiments to study the binding location of small-molecules and ssDNA-based inhibitors to $A3B_{CTD}$. We decided to utilise SAXS to examine the relative structure and structural features of our active $A3B_{CTD}$-AL1 with our dZ-modified ssDNA inhibitor. We obtained the first low-resolution SAXS models of a catalytically active $A3B_{CTD}$ in complex with the dZ-containing ssDNA inhibitor. We determined that the ssDNA-based inhibitor binds to the active $A3B_{CTD}$ in a similar manner as that reported for the inactive $A3B_{CTD}$ in complex with substrate ssDNA (PDB, 5TD5)

[239]. This further supports' our proposed competitive inhibition mechanism of dZ-containing DNA (refer to Figure 3.10).

Moreover, our $A3B_{CTD}$-AL1 also eluted as a larger oligomeric species. Using a combination interface modelling and SAXS we confirmed that this species was an $A3B_{CTD}$-AL1 dimer in solution that was formed through several weak interactions and stabilised through a disulfide bond. It was apparent that this construct dimerised as a result of the loop 1 from the A3A, likely coincidentally similar to a previously reported A3A crystallographic dimer (PDB, 4xxo [23]). As described in several studies [22-23, 27, 79, 118, 233], the potential of oligomerisation of A3 enzymes may regulate the catalytic activity of the enzymes. Oligomerisation of our construct may allow added stability and inactivate the catalytic activity. We can deduce that dimerisation may mimic a regulatory mechanism of the A3A enzyme; however, whether this occurs in the cell remains unknown.

To model the structure of two-domain A3 enzymes, the full-length two-domain A3G was used. In solution the ligand-free full-length A3G eluted off the SEC column in several conformational states, but largely as high molecular weight (HMW) oligomers, consistent with several studies [48, 50, 55]. These HMW A3G species are reported to be formed through non-specific interactions with RNA leading to their inactivation [23, 48, 50, 55, 150, 276], but these species can be reversed to active low molecular weight (LMW) A3G forms upon treatment with RNase A [48, 55]. In our study, the A3G tetramer, made up of four two-domain A3G subunits, was predominant species in solution. However, we could not determine the interface and whether our A3G tetramer was bound through RNA. Upon the addition of dZ-containing ssDNA, our A3G tetramer forms a complex with two ssDNA molecules, and triggers disassembly into LMW A3G conformations (A3G dimer and monomer), rather than by the treatment with RNase as indicated in previous studies [48, 55]. We predict that the RNA may bind non-specifically near the active-site of the two-domain A3 causing oligomerisation and hindrance of the enzyme's catalytic activity. However, upon binding of ssDNA to HMW conformations the enzyme dismantles into the active LMW states, supporting a catalytic regulatory mechanism. Higher resolution structures would help elucidate this structure and understand the mechanism of regulation. Notably, we report the first SAXS model structure of the two-domain A3G in complex with a ssDNA inhibitor, which has been largely unknown. From our in-solution model, we can deduce that the ssDNA-inhibitor binds into the CTD, whereby dZ binds with a high affinity into the active site of the CTD. Somewhat unexpectedly the rest of

the oligonucleotide remained flexible with limited interactions with the NTD (Figure 6.21), inconsistent with studies implying that the NTD is required in binding ssDNA to enhance deamination at the CTD [24, 78, 88, 91, 187, 189, 247]. This suggests that the mechanism of ssDNA binding to two-domain A3 must be highly dynamic and needs to be modelled. Overall, this work provides a platform for the development of selective A3B-inhibitors and further structural studies with modified ssDNA inhibitors. We aim to develop highly potent selective A3B ssDNA-based inhibitors and evaluate them in cellular studies and ultimately to be utilised as possible conjuvants in anti-cancer treatments.

7.2 Future directions

7.2.1 Structural studies

Using the reported full-length A3G dimer structures [286], we can use these structures to determine the interface between two molecules (PRISM 2.0 webserver [9, 264]) and build low resolution models of our ligand-free A3G tetramer. This would advance our understanding of the interactions that drive oligomerisation of A3G and other A3 enzymes, as indicated in the literature [22-23, 27, 79, 118, 233].

Using our $A3B_{CTD}$-QM-ΔL3 variant assignment, we can use chemical shift perturbation experiments to further examine the interactions between the $A3B_{CTD}$ and our ssDNA-based inhibitors to better understand the binding mechanism. In addition, X-ray crystallography can be utilised to determine the structures of the active $A3B_{CTD}$ and the multiple forms of two-domain A3 structures in complex with our ssDNA-based inhibitors.

7.2.2 Evaluation of ssDNA-based inhibitors using wild-type A3 enzymes

In our experiments, we used $A3B_{CTD}$ constructs that provided the starting point in the development of ssDNA-based inhibitors. We noted that the inhibition effect of our ssDNA-based inhibitors on the full-length two domain A3G was similar to the inhibition of the single catalytic domain $A3G_{CTD}$ [136], as deaminase activity is largely attributed to the CTD. This indicates that the two-domain A3B would equally be inhibited in a similar manner as that detected for the $A3B_{CTD}$. However, to verify this presumption we would need to evaluate our ssDNA-based inhibitors on full-length A3B.

7.2.3 A3B inhibitors with enhanced potencies

Future work will focus on optimisations of our substrate-like ssDNA inhibitors. The use of more potent CDA cytosine analogue nucleoside inhibitors [6, 45, 277] incorporated into the A3B preferred ssDNA sequences, in principle will enhance the inhibition potential of these oligonucleotides. Furthermore, as noted in our study, nucleotides flanking the target dZ influence the binding specificity of the A3 enzymes, as these enzymes have different electrostatic and hydrophobic distribution around the active site. Therefore, modifications of these flanking regions by placement of large aromatic moieties may improve inhibitory potential, perhaps by acting as molecular caps closing the ssDNA binding site and stabilising the complex by removing water molecules. Moreover, as indicated by our SAXS models and inhibition mechanism, our modified-ssDNA inhibitors appear to bind to the $A3B_{CTD}$ in a similar manner as substrate ssDNA [239] (PDB, 5CQI), which conforms in a U-shape. Therefore, our linear modified-ssDNAs must undergo several conformational changes to stabilise this form and position the cytosine analogue nucleoside into the active-site of the enzyme. So, by pre-shaping our modified-ssDNA inhibitors in this U-shaped conformation we assume that the binding affinity to the enzyme will be improved, thereby boosting the inhibition potential of these inhibitors. Evaluation of ssDNA with more potent transition-state analogues of cytosine deamination are expected to increase inhibition to the level required for the cellular studies. These inhibitors can then be used as evaluated with existing anticancer therapies in cancer cells, animals' models and if successful in humans.

8. Appendices

8.1 Supplementary figures

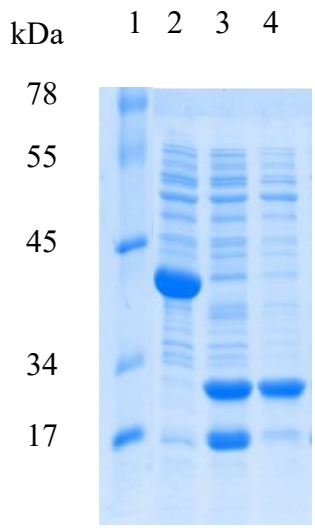


Figure 8.1 Cleavage of SUMO fusion tag from His-SUMO-A3B$_{CTD}$-DM using SUMO protease.
Cleavage process as described in section 2.2.5.1. Tricine 16% SDS PAGE gel, Lane 1; SeeBlue Plus2 Pre-stained protein standard ladder, Lane 2; SUMO-A3B$_{CTD}$-DM, Lane 3; His$_6$-SUMO-A3B$_{CTD}$-DM treated with SUMO protease at 1:5 ratio, Lane 4; A3B$_{CTD}$-DM eluted off IMAC column.

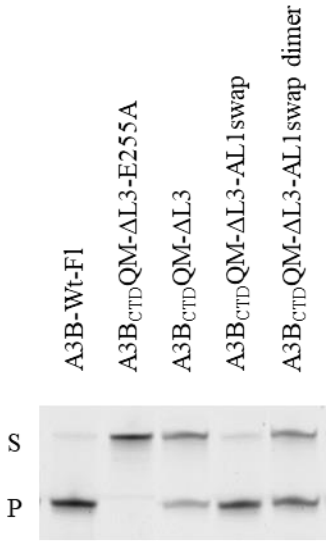


Figure 8.2 *In vitro* deaminase activity assay of the A3B$_{CTD}$ variants.
An established *in vitro* DNA deamination in-gel based assay [240] was performed using a final concentration of 5 µM of our purified A3B variants along with 800 nM of a fluorescein tagged oligonucleotide (TC 3' 6-FAM) as described in section 2.5.1. The wild type full-length A3B (A3B-Wt-Fl) protein was used as a control of catalytic activity.

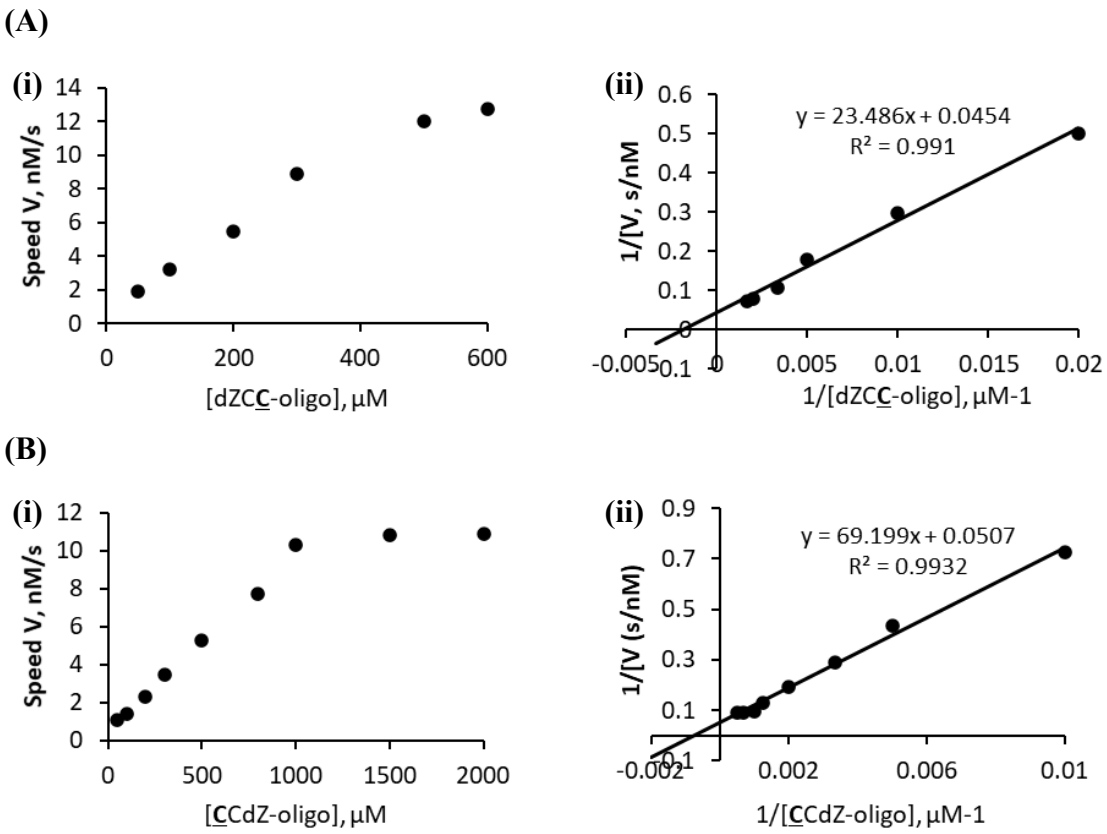


Figure 8.3 Deamination speed of the A3 enzymes on the preferred dC in dZ containing CCC-motifs.

Deamination speed measured using ^{1}H-NMR at 298 K of (A) A3A-mimic (50 nM) as a function of dZCC-oligo substrate concentration in 50 mM citrate-phosphate buffer at pH 5.5, (B) $A3B_{CTD}$-DM (2 µM) as a function of CCdZ-oligo substrate concentration 50 mM citrate-phosphate buffer at pH 7.5. Initial speed of deamination obtained at varying substrate concentrations (2 – 2500 µM) (i) Plot of the speed of deamination as a function of substrate concentration, (ii) Lineweaver-Burk plot of the inverse speed of deamination as a function of inverse substrate concentration. In all NMR assays, buffers contained 10% D_2O.

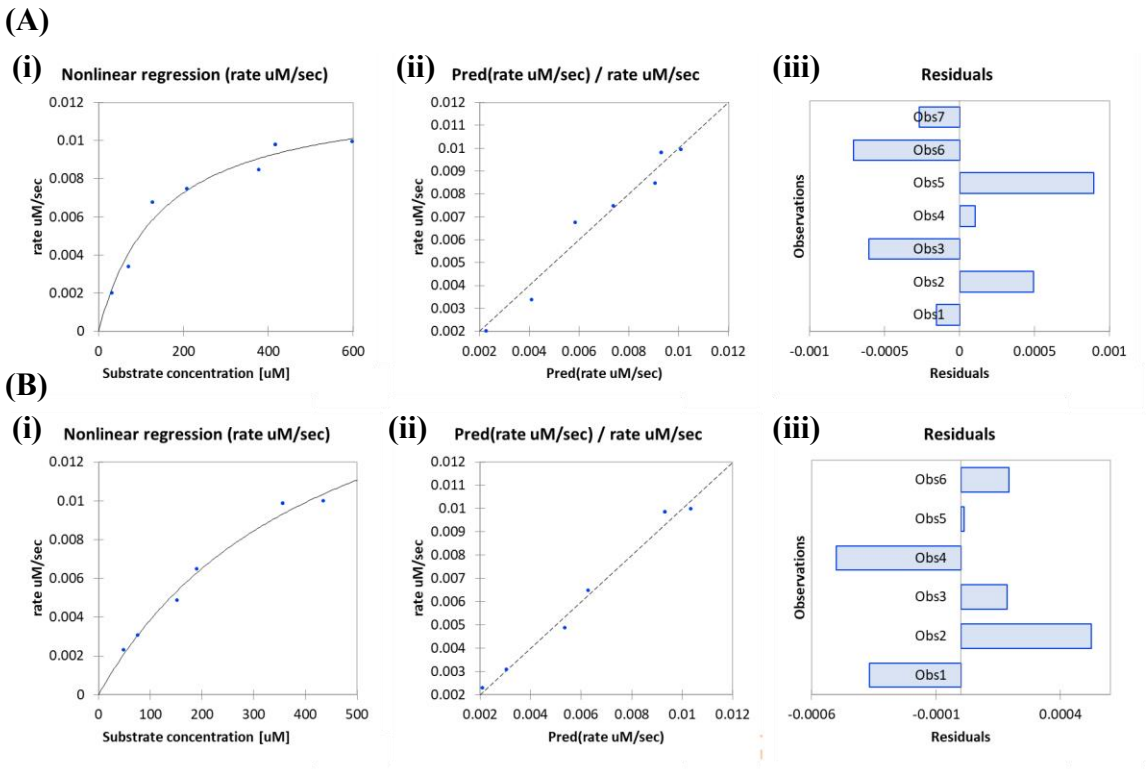


Figure 8.4 Non-linear regression analysis of A3B$_{CTD}$-AL1 and A3B$_{CTD}$-DM cytosine deamination of 5'-ATTTCATTT.

Non-linear regression analysis of the rate of cytosine deamination of the 5'-TCA-oligo at various substrate concentration by (A) A3B$_{CTD}$-AL1 or (B) A3B$_{CTD}$-DM. (i) Plot of initial rate of deamination as a function of substrate concentration fitted with a non-linear regression model to derive K_m and k_{cat} parameters, (ii) predicted values from model from each observation plotted against the observed data (iii) residuals calculated from the model for each observation. The derived parameters are tabulated in Table 8.3. I acknowledge Prof. Geoffrey B. Jameson for conducting non-linear regression analysis on derived data using excel XLSAT addon program.

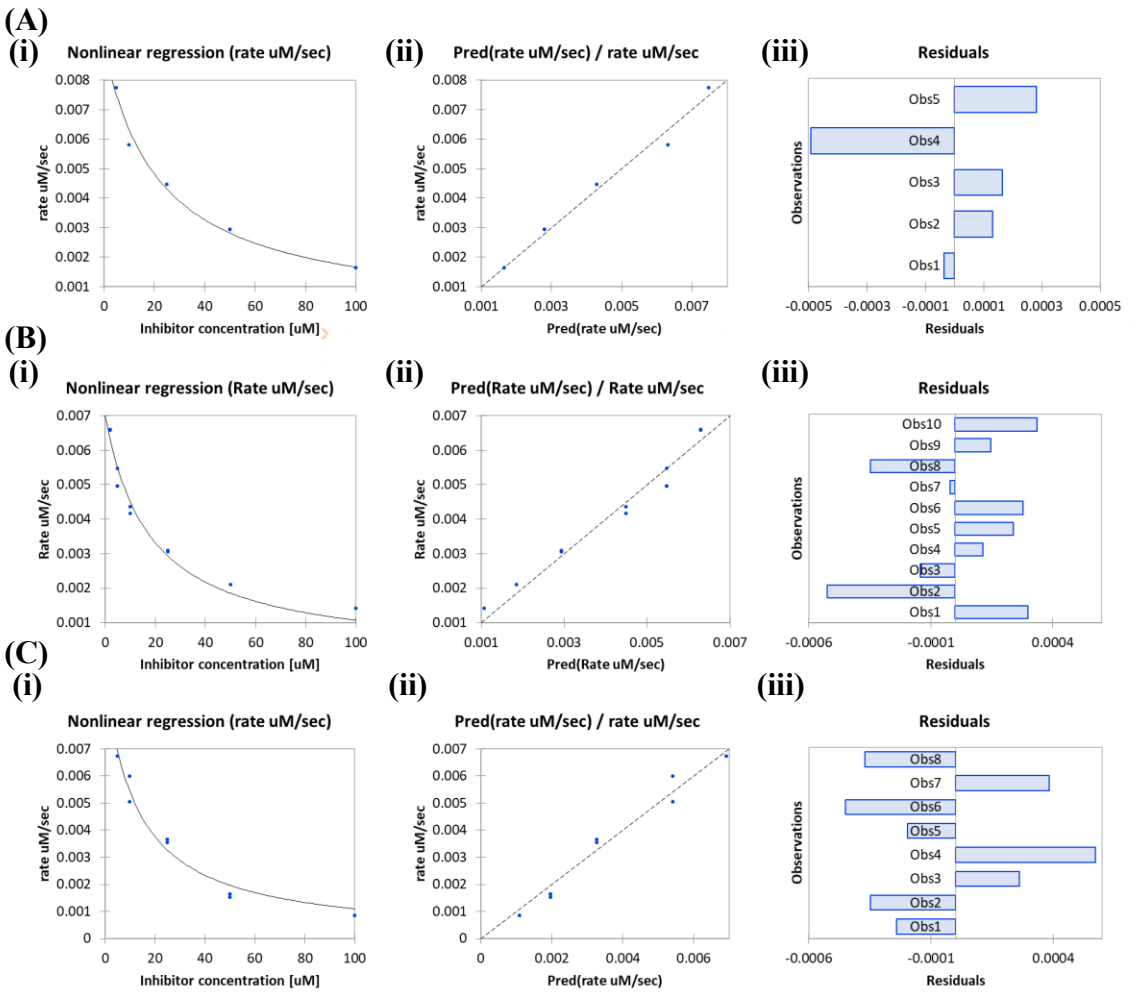


Figure 8.5 Non-linear regression analysis of the inhibition of A3B$_{CTD}$-AL1 and A3B$_{CTD}$-DM cytosine deamination by dZ- and 5FdZ-containing oligos.

Non-linear regression analysis of the rate of cytosine deamination of the 5'-TCA-oligo by A3B$_{CTD}$-variants in the presence dZ- and 5FdZ-containing oligos. (A) A3B$_{CTD}$-AL1 in the presence of dZ-oligo, (B) A3B$_{CTD}$-DM in the presence of dZ-oligo, and (C) A3B$_{CTD}$-AL1 in the presence of 5FdZ-oligo. (i) Plot of initial rate of deamination as a function of inhibitor concentration fitted with a non-linear regression model to derive K_i, (ii) predicted values from model from each observation plotted against the observed data (iii) residuals calculated from the model for each observation. The derived parameters are tabulated in Table 8.4. I acknowledge Prof. Geoffrey B. Jameson for conducting non-linear regression analysis on derived data using excel XLSAT addon program.

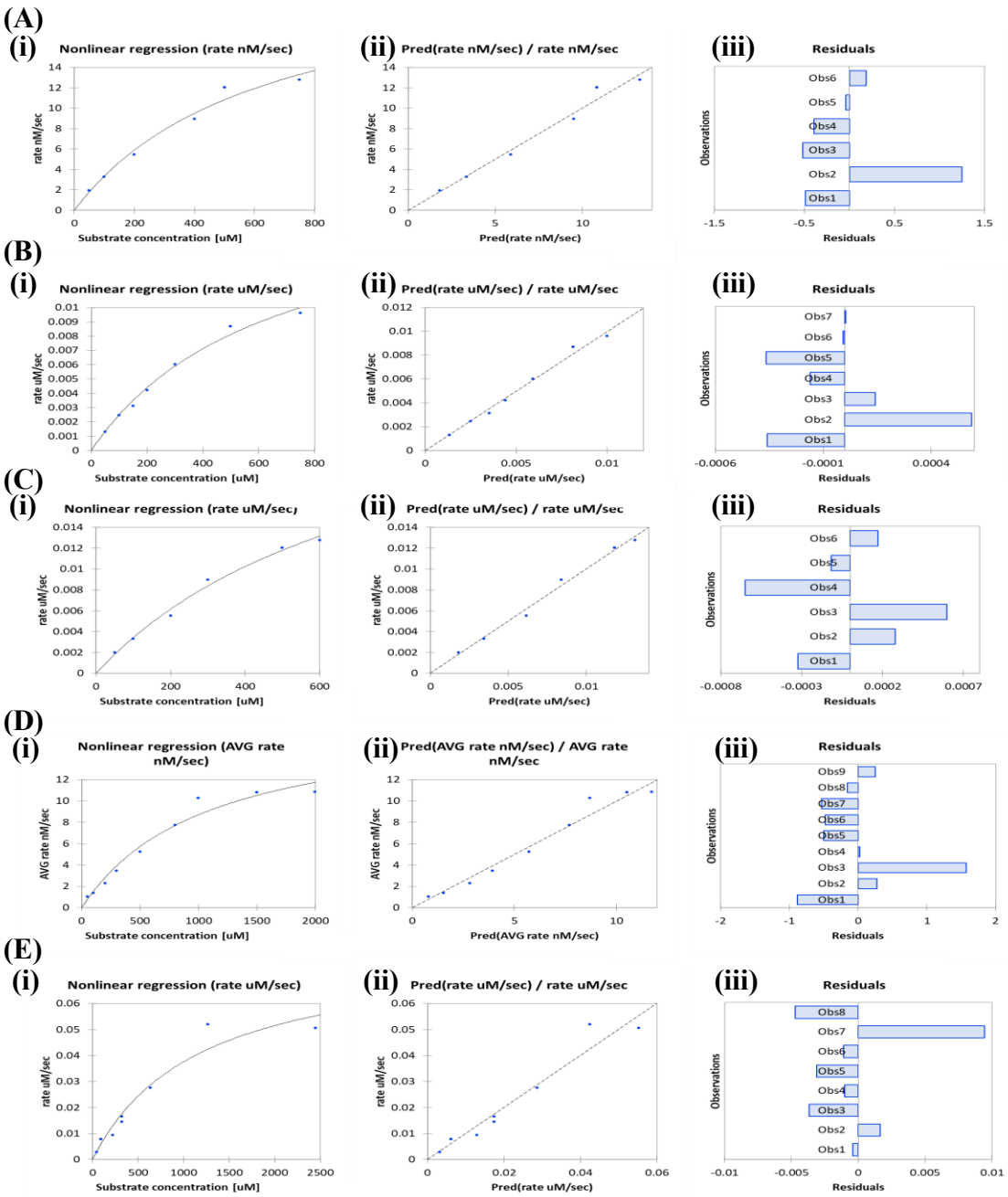


Figure 8.6 Non-linear regression analysis of A3A-mimic, A3B$_{CTD}$-DM, A3G$_{CTD}$-Wt on the deamination of their selective dC.

Non-linear regression analysis of the rate of cytosine deamination of by (A) A3A-mimic deamination of dC near the 3'-end in the 5'-CC<u>C</u>-oligo, (B) A3B$_{CTD}$-DM deamination of dC near the 5'-end in the 5'-<u>C</u>CC-oligo, (C) A3A-mimic deamination of dC near the 3'-end in the 5'-dZC<u>C</u>-oligo, (D) A3B$_{CTD}$-DM deamination of dC near the 5'-end in the 5'-<u>C</u>CdZ-oligo, (E) A3G$_{CTD}$-Wt deamination of dC near the 3'-end in the 5'-dZC<u>C</u>-oligo. (i) Plot of initial rate of deamination as a function of substrate concentration fitted with a non-linear regression model to derive K_m and k_{cat} parameters, (ii) predicted values from model from each observation plotted against the observed data (iii) residuals calculated from the model for each observation. The derived parameters are tabulated in Table 8.3. I acknowledge Prof. Geoffrey B. Jameson for conducting non-linear regression analysis on derived data using excel XLSAT addon program.

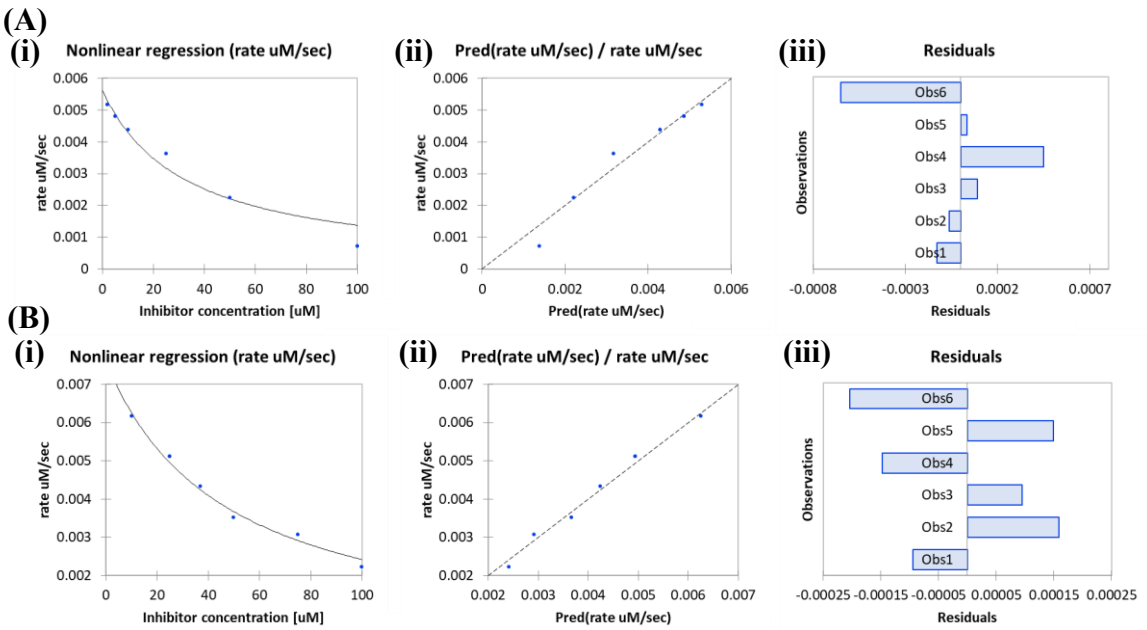


Figure 8.7 Non-linear regression analysis of the inhibition of A3A-mimic and A3B$_{CTD}$-DM cytosine deamination by selective dZ-containing oligos.

Non-linear regression analysis of the rate of cytosine deamination of the 5'-TCA-oligo by A3B$_{CTD}$-variants in the presence selective dZ-containing oligos. (A) A3A-mimic dC deamination of the TCA-oligo in the presence of the CCdZ-oligo, (B) A3B$_{CTD}$-DM dC deamination of the TCA-oligo in the presence of the dZCC-oligo. (i) Plot of initial rate of deamination as a function of inhibitor concentration fitted with a non-linear regression model to derive K_i, (ii) predicted values from model from each observation plotted against the observed data (iii) residuals calculated from the model for each observation. The derived parameters are tabulated in Table 8.4. I acknowledge Prof. Geoffrey B. Jameson for conducting non-linear regression analysis on derived data using excel XLSAT addon program.

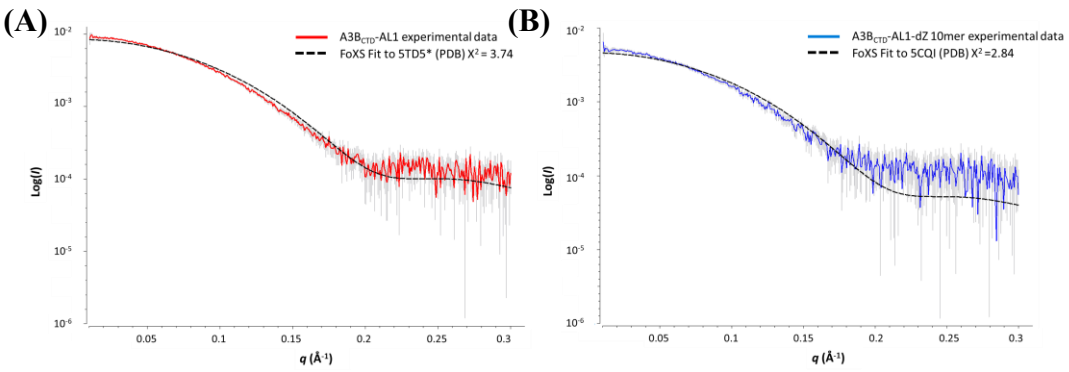


Figure 8.8 Fitting of A3B$_{CTD}$-AL1 and A3B$_{CTD}$-AL1-dZ-oligo SAXS profiles to A3B$_{CTD}$-DNA (5TD5*) and A3B$_{CTD}$ (5CQI) crystal structures, respectively.

Model fit of 5TD5* to A3B$_{CTD}$-AL1 scattering data (left), fit of 5CQI to A3B$_{CTD}$-AL1-dZ oligo (1 to 2 ratio) scattering data.

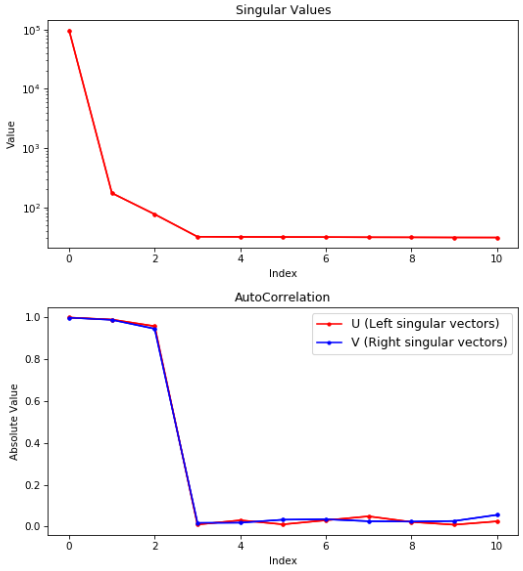


Figure 8.9 Singular value decomposition (SVD) of A3B$_{CTD}$-AL1 dimer with dZ-oligo sample. SAXS profile assessed using SVD/EFA BioXTAS RAW [105]. Top panel represents SVD showing the Eigen value on the index represents number of eigenvalues equivalent to number of components in scattering sample. Top panel represents SVD showing an eigenvalue of three. Bottom panel displays the autocorrelation between the singular vectors (blue and red lines). indicating that the singular values are not so variable.

```
A3B           1 MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVY
A3G           1 MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTK-GPSRPPLDAKIFRGQVY
                * *  **  ********  ** * ***  *   ******* * * *    *   ******

A3B          61 FKPQYHAEMCFLSWFCG-NQLPAYKCFQITWFVSWTPCPDCVAKLAEFLSEHPNVTLTIS
A3G          60 SELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIF
                ** ** *  **     *         **  ** **  *    * ** * * *****

A3B         120 AARLYYYWERDYRRALCRLSQA--GARVT--IMDYEEFAYCWENFVYNEGQQFMPWYKFD
A3G         120 VARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLP
               ***** *  **  **  * *   * * *  ** * **  **  ***     * **

A3B         176 ENYAFLHRTLKEILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTWVLMDQHM
A3G         180 KYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRR
              *  **  * ****  *** ******** * *  *  *********  * ****  *

A3B         236 GFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRA
A3G         240 GFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFS--CAQEMAK
                ***** *        ****** ***  *   **  * **** * *******  ** *

A3B         296 FLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCP
A3G         298 FISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCP
                *   * ** * ** *****      * *  *  ***  ***** **  ******  ****

A3B         356 FQPWDGLEEHSQALSGRLRAILQNQGN
A3G         358 FQPWDGLDEHSQDLSGRLRAILQNQEN
                ******* **** ************ *
```

Figure 8.10 Comparison between the amino acid sequence of A3B and A3G.

* represents the same amino acid

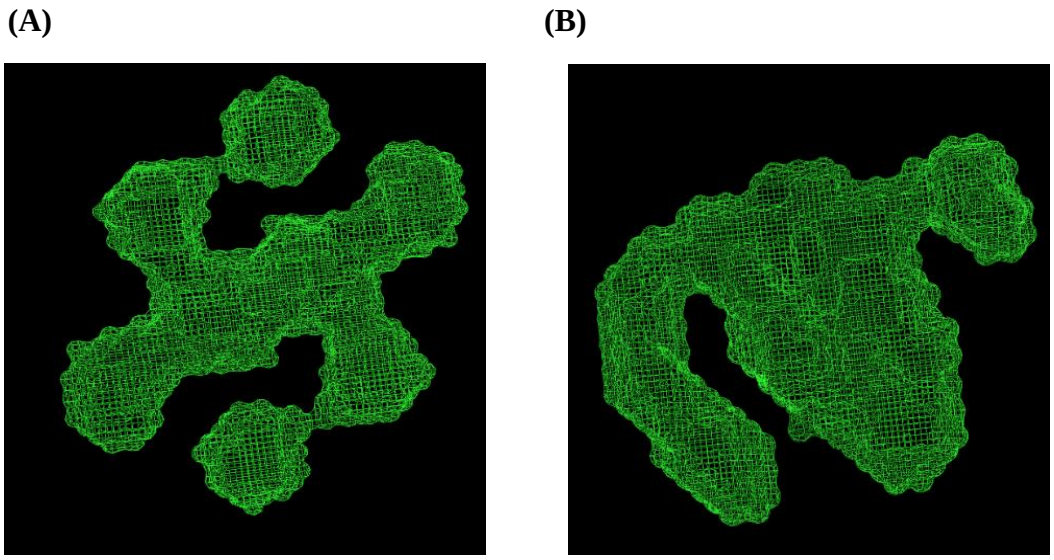


Figure 8.11 Envelope models of A3B$_{CTD}$-AL1 dimer/dZ-oligo scattering species.
(A) Average dummy-model of species 1, an A3B$_{CTD}$-AL1 dimer, generated using *P*2 symmetry with an NSD of 0.993, (B) Average dummy-model of species 2, an A3B$_{CTD}$-AL1 monomer, generated using *P*1 symmetry with an NSD of 1.006. For each species ten models were generated using DAMMIF, then filtered using DAMAVER and DAMFILT (ATSAS 2.8.3 suite) [75]

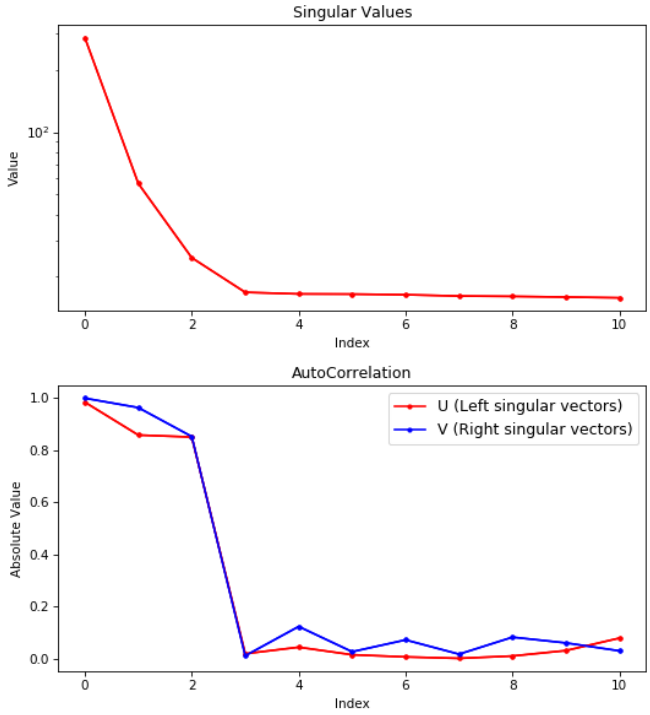


Figure 8.12 Singular value decomposition (SVD) of A3Gfl-Wt sample.
SAXS profile assessed using SVD/EFA BioXTAS RAW [105], index represents number of eigenvalues equivalent to number of components in scattering sample. Top panel represents SVD showing an eigenvalue of three. Bottom panel displays the autocorrelation between the singular vectors (blue and red lines). indicating that the singular values are not so variable.

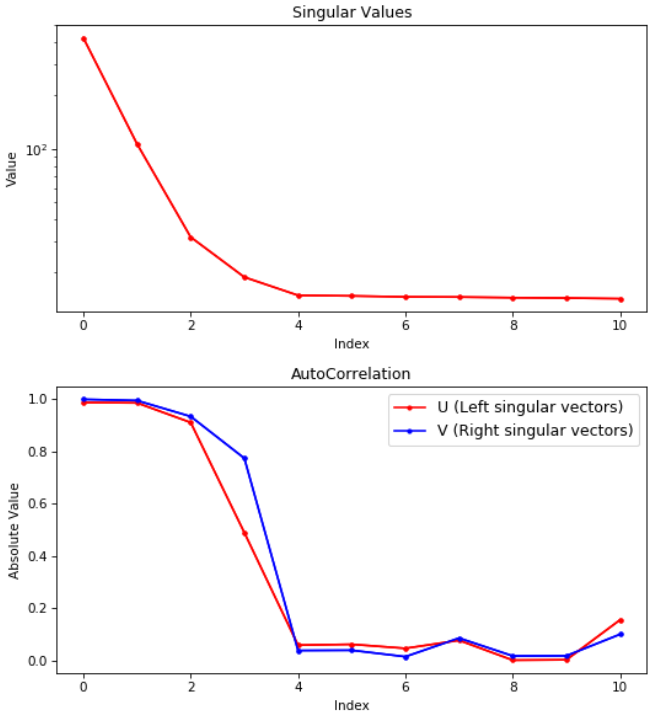


Figure 8.13 Singular value decomposition (SVD) of A3Gfl-Wt with CCdZ-40mer oligo sample. SAXS profile assessed using SVD/EFA BioXTAS RAW [105]. Top panel represents SVD showing the Eigen value on the index represents number of eigen values equivalent to number of components in scattering sample. Top panel represents SVD showing an Eigen value of four. Bottom panel displays the autocorrelation between the singular vectors (blue and red lines). indicating that the singular values are not too variable.

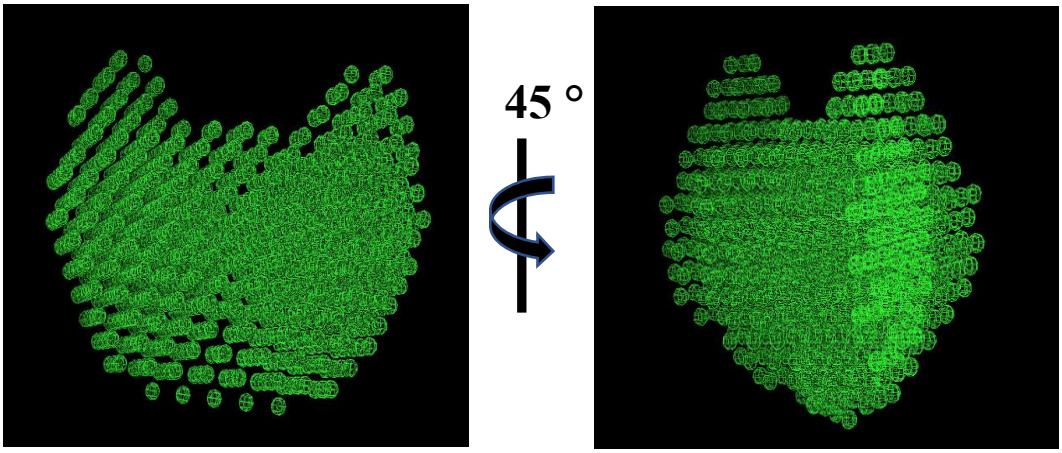


Figure 8.14 A3Gfl with 40mer ssDNA species 3 (A3G dimer) envelope model. Average dummy-model generated using *P*2 symmetry with an NSD of 0.937. For each species ten models were generated using DAMMIF, then filtered using DAMAVER and DAMFILT (ATSAS 2.8.3 suite) [75].

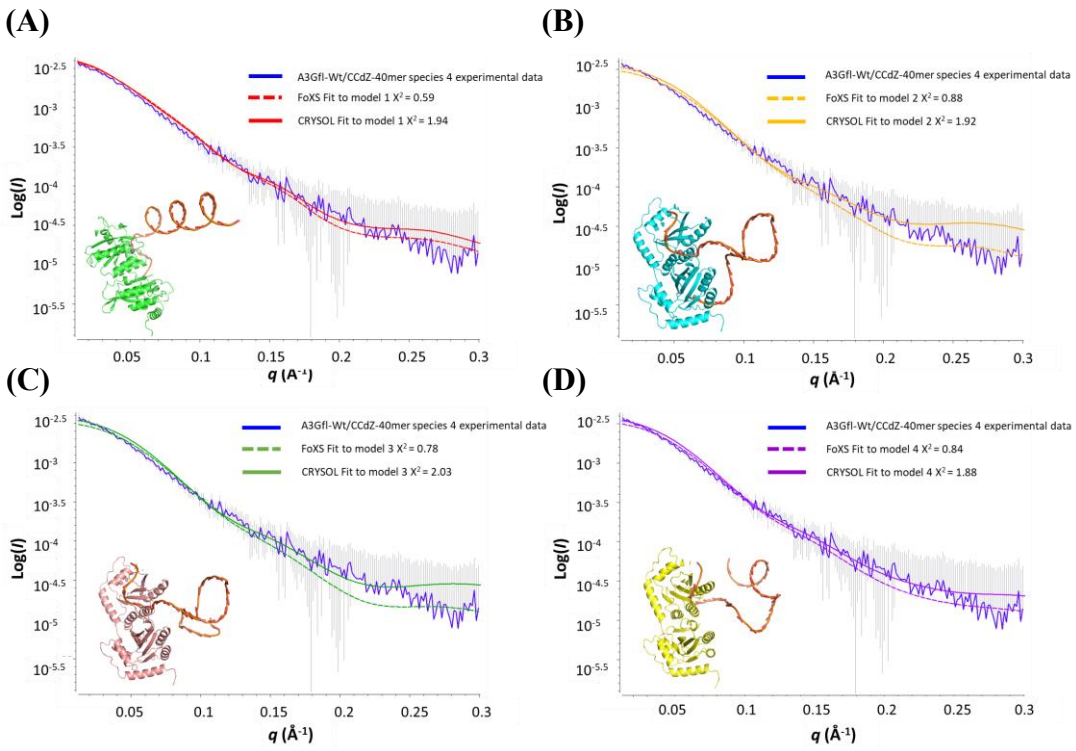


Figure 8.15 Rigid body modelling of A3G monomer with CCdZ-40mer (species 4) scattering data to envelope generated hypothetical models.

Fitting DNA bound A3G monomer models (inserts) (A) model 1, (B) model 2, (C) model 3, (D) model 4 into A3G monomer 40mer ssDNA scattering data.

8.2 Supplementary equations

$$b = \frac{1}{k_{cat}[E]}$$

$$k_{cat} = \frac{1}{b \times [E]}$$

$$\frac{\partial k_{cat}}{\partial b} = -\frac{1}{b^2} \times \frac{1}{[E]}$$

$$\sigma^2 k_{cat} = \left(\frac{1}{b^2} \times \frac{1}{[E]}\right)^2 \times \sigma^2(b)$$

$$\sigma k_{cat} = \sqrt{\sigma^2 k_{cat}}$$

$$(\boldsymbol{i})\ \sigma^2 K_m = (k_{cat} \times [E])^2 \times \sigma^2(a)$$

$$\frac{\partial K_m}{\partial k_{cat}} = a \times [E]$$

$$(\boldsymbol{ii})\ \sigma^2 K_m = (a \times [E])^2 \times \sigma^2(k_{cat})$$

$$\sigma^2 K_m = (\boldsymbol{i}) + (\boldsymbol{ii})$$

$$\sigma K_m = \sqrt{\sigma^2 K_m}$$

$$a = \frac{K_m}{k_{cat}[E]}$$

$$K_m = a \times k_{cat} \times [E]$$

$$\frac{\partial K_m}{\partial a} = k_{cat} \times [E]$$

Equation 8.3 Calculation of uncertainties for K_m and k_{cat} using error propagation method.

Errors of linear regression fit to measured data were calculated using LINEST function in Excel (termed σ*a* and σ*b* in equation). Uncertainties of K_m and k_{cat} (termed σK_m and σk_{cat} in equation) calculated using error-propagation method by using partial derivatives. I acknowledge the assistance of Prof. Geoffrey B. Jameson in deriving equations for error propagation.

$$a = \frac{K_m}{K_i \times V_{max} \times [S]}$$

$$a = \frac{K_m}{K_i} \times \frac{1}{V_{max}} \times \frac{1}{[S]}$$

$$K_i = \frac{K_m}{a} \times \frac{1}{V_{max}} \times \frac{1}{[S]}$$

$$\frac{\partial K_i}{\partial a} = -\frac{1}{a^2} \times K_m \times \frac{1}{V_{max}} \times \frac{1}{[S]}$$

$$b = \frac{K_m + [S]}{V_{max} \times [S]}$$

$$\mathbf{i})\ \sigma^2 K_i = \left(\frac{K_m}{a^2[S]} \times \frac{1}{V_{max}}\right)^2 \times \sigma^2(a)$$

$$\frac{\partial K_i}{\partial K_m} = \frac{1}{a} \times \frac{1}{V_{max}} \times \frac{1}{[S]}$$

$$\mathbf{ii})\ \sigma^2 K_i = \left(\frac{1}{a} \times \frac{1}{V_{max}} \times \frac{1}{[S]}\right)^2 \times \sigma^2(K_m)$$

$$\frac{\partial K_i}{\partial(\frac{1}{V_{max}})} = \frac{1}{a} \times K_m \times \frac{1}{[S]}$$

$$\mathbf{iii})\ \sigma^2 K_i = \left(\frac{1}{a} \times K_m \times \frac{1}{[S]}\right)^2 \times \sigma^2\left(\frac{1}{V_{max}}\right)$$

$$\sigma^2 K_i = \boldsymbol{i}) + \boldsymbol{ii}) + \boldsymbol{iii})$$

$$\sigma K_i = \sqrt{\sigma^2 K_i}$$

Equation 8.4 Calculation of uncertainty for K_i using error propagation method.

Errors of linear regression fit to measured data were calculated using LINEST function in Excel (termed σ*a* and σ*b* in equation). Uncertainty of K_i (termed σK_i in equation) calculated using error-propagation method by using partial derivatives. I acknowledge the assistance of Prof. Geoffrey B. Jameson in deriving equations for error propagation.

8.3 Supplementary tables

Table 8.1 Correlations of Multidimensional NMR experiments

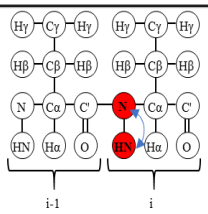
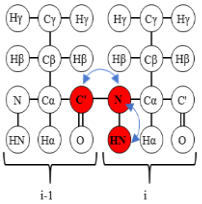
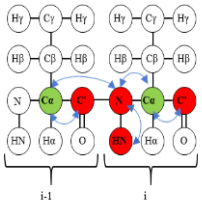
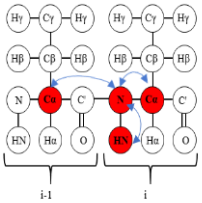
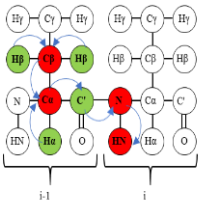
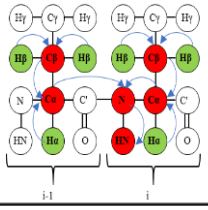
NMR experiments	Nuclei correlations	
^{1}H-^{15}N HSQC	Correlation between the ^{15}N nuclei and amide proton of the residue (i).	
HNCO	Correlation between the amide chemical shift with the carbonyl carbon of the previous residue (i-1).	
HN(CA)CO	Correlation between the amide chemical shifts through the alpha carbons (Cα) (of the residue (i) and the previous residue (i-1)), with the carbonyl carbons of the residue (i) and the previous residue (i-1).	
HNCA	Correlation between the amide chemical shifts with the alpha carbon (Cα) of the residue (i) and the previous residue (i-1).	
CBCA(CO)NH	Correlation between the amide chemical shifts of the residue (i) through the carbonyl carbons with both Cα and Cβ of the previous residue (i-1).	
HNCACB	Correlation of the chemical shift of amide to the Cα and Cβ of the residue (i) and the preceding residue (i-1).	

Table 8.2 NMR assignment of A3B$_{CTD}$-QM-ΔL3

Atom shift assignment ID	Residue sequence code	Residue label	Atom name	Atom type	Chemical shift value (ppm)	Chemical shift value error (ppm)	Chemical shift ambiguity code
1	7	GLU	CA	C	55.488	0.3	1
2	7	GLU	H	H	7.721	0.02	1
3	7	GLU	HA	H	3.87	0.02	1
4	7	GLU	N	N	118.399	0.3	1
5	8	ILE	CA	C	53.919	0.3	1
6	8	ILE	H	H	7.992	0.02	1
7	8	ILE	N	N	122.086	0.3	1
8	9	LEU	H	H	8.428	0.02	1
9	9	LEU	N	N	120.968	0.3	1
10	11	TYR	CA	C	59.942	0.3	1
11	11	TYR	H	H	8.201	0.02	1
12	11	TYR	N	N	122.284	0.3	1
13	12	LEU	CA	C	53.974	0.3	1
14	12	LEU	H	H	8.125	0.02	1
15	12	LEU	N	N	126.184	0.3	1
16	13	MET	CA	C	54.84	0.3	1
17	13	MET	H	H	8.051	0.02	1
18	13	MET	HA	H	4.522	0.02	1
19	13	MET	N	N	120.868	0.3	1
20	14	ASP	CA	C	56.483	0.3	1
21	14	ASP	H	H	8.031	0.02	1
22	14	ASP	HA	H	3.829	0.02	1
23	14	ASP	N	N	120.568	0.3	1
24	16	ASP	CA	C	56.083	0.3	1
25	16	ASP	H	H	8.409	0.02	1
26	16	ASP	HA	H	4.158	0.02	1
27	16	ASP	N	N	117.476	0.3	1
28	17	THR	CA	C	65.49	0.3	1
29	17	THR	H	H	7.931	0.02	1
30	17	THR	N	N	120.047	0.3	1
31	18	PHE	CA	C	61.626	0.3	1
32	18	PHE	H	H	7.935	0.02	1
33	18	PHE	N	N	121.935	0.3	1
34	19	THR	CA	C	65.599	0.3	1
35	19	THR	H	H	9.156	0.02	1
36	19	THR	HA	H	4.685	0.02	1
37	19	THR	N	N	116.152	0.3	1
38	20	SER	CA	C	59.857	0.3	1
39	20	SER	H	H	8.051	0.02	1
40	20	SER	HA	H	4.062	0.02	1
41	20	SER	N	N	113.364	0.3	1
42	21	ASN	CA	C	54.987	0.3	1
43	21	ASN	H	H	7.226	0.02	1
44	21	ASN	HA	H	4.171	0.02	1
45	21	ASN	N	N	114.551	0.3	1
46	23	ASN	CA	C	53.547	0.3	1
47	23	ASN	H	H	5.882	0.02	1
48	23	ASN	HA	H	4.529	0.02	1
49	23	ASN	N	N	114.573	0.3	1
50	24	ASN	CA	C	50.591	0.3	1
51	24	ASN	H	H	6.833	0.02	1
52	24	ASN	HA	H	4.433	0.02	1
53	24	ASN	N	N	123.512	0.3	1
54	27	LEU	CA	C	53.763	0.3	1

Atom shift assignment ID	Residue sequence code	Residue label	Atom name	Atom type	Chemical shift value (ppm)	Chemical shift value error (ppm)	Chemical shift ambiguity code
55	27	LEU	H	H	8.46	0.02	1
56	27	LEU	HA	H	4.125	0.02	1
57	27	LEU	N	N	118.772	0.3	1
58	28	VAL	CA	C	61.519	0.3	1
59	28	VAL	H	H	7.006	0.02	1
60	28	VAL	HA	H	4.174	0.02	1
61	28	VAL	N	N	120.103	0.3	1
62	29	LEU	CA	C	54.25	0.3	1
63	29	LEU	H	H	8.2	0.02	1
64	29	LEU	N	N	127.47	0.3	1
65	30	ARG	CA	C	52.676	0.3	1
66	30	ARG	H	H	8.548	0.02	1
67	30	ARG	HA	H	4.736	0.02	1
68	30	ARG	N	N	126.729	0.3	1
69	31	ARG	CA	C	58.763	0.3	1
70	31	ARG	H	H	7.898	0.02	1
71	31	ARG	N	N	123.848	0.3	1
72	35	TYR	CA	C	56.67	0.3	1
73	35	TYR	H	H	8.362	0.02	1
74	35	TYR	N	N	126.124	0.3	1
75	36	LEU	CA	C	53.22	0.3	1
76	36	LEU	H	H	9.131	0.02	1
77	36	LEU	HA	H	4.782	0.02	1
78	36	LEU	N	N	126.345	0.3	1
79	37	CYS	CA	C	56.179	0.3	1
80	37	CYS	H	H	9.618	0.02	1
81	37	CYS	HA	H	5.006	0.02	1
82	37	CYS	N	N	126.063	0.3	1
83	38	TYR	CA	C	53.481	0.3	1
84	38	TYR	H	H	8.955	0.02	1
85	38	TYR	HA	H	4.038	0.02	1
86	38	TYR	N	N	121.133	0.3	1
87	39	GLU	CA	C	53.926	0.3	1
88	39	GLU	H	H	8.623	0.02	1
89	39	GLU	HA	H	3.657	0.02	1
90	39	GLU	N	N	113.776	0.3	1
91	40	VAL	CA	C	60.811	0.3	1
92	40	VAL	H	H	8.623	0.02	1
93	40	VAL	N	N	120.972	0.3	1
94	41	GLU	CA	C	52.837	0.3	1
95	41	GLU	H	H	9.157	0.02	1
96	41	GLU	HA	H	5.666	0.02	1
97	41	GLU	N	N	124.882	0.3	1
98	42	ARG	CA	C	54.171	0.3	1
99	42	ARG	H	H	9.154	0.02	1
100	42	ARG	HA	H	5.365	0.02	1
101	42	ARG	N	N	124.609	0.3	1
102	43	LEU	CA	C	51.88	0.3	1
103	43	LEU	H	H	8.47	0.02	1
104	43	LEU	HA	H	4.3	0.02	1
105	43	LEU	N	N	124.654	0.3	1
106	44	ASP	CA	C	53.555	0.3	1
107	44	ASP	H	H	8.015	0.02	1
108	44	ASP	N	N	127.028	0.3	1
109	45	ASN	CA	C	54.839	0.3	1
110	45	ASN	H	H	8.753	0.02	1

Atom shift assignment ID	Residue sequence code	Residue label	Atom name	Atom type	Chemical shift value (ppm)	Chemical shift value error (ppm)	Chemical shift ambiguity code
111	45	ASN	N	N	121.739	0.3	1
112	46	GLY	CA	C	44.449	0.3	1
113	46	GLY	H	H	8.79	0.02	1
114	46	GLY	HA2	H	3.997	0.02	1
115	46	GLY	N	N	111.175	0.3	1
116	47	THR	CA	C	59.9	0.3	1
117	47	THR	H	H	7.739	0.02	1
118	47	THR	HA	H	4.524	0.02	1
119	47	THR	N	N	116.314	0.3	1
120	48	SER	CA	C	56.455	0.3	1
121	48	SER	H	H	8.546	0.02	1
122	48	SER	N	N	119.457	0.3	1
123	49	VAL	CA	C	60.317	0.3	1
124	49	VAL	H	H	9.02	0.02	1
125	49	VAL	HA	H	5.218	0.02	1
126	49	VAL	N	N	125.331	0.3	1
127	50	LYS	CA	C	56.374	0.3	1
128	50	LYS	H	H	8.766	0.02	1
129	50	LYS	HA	H	4.285	0.02	1
130	50	LYS	N	N	127.436	0.3	1
131	51	MET	CA	C	53.734	0.3	1
132	51	MET	H	H	8.526	0.02	1
133	51	MET	HA	H	5.147	0.02	1
134	51	MET	N	N	124.674	0.3	1
135	52	ASP	CA	C	54.302	0.3	1
136	52	ASP	H	H	8.381	0.02	1
137	52	ASP	HA	H	3.886	0.02	1
138	52	ASP	N	N	123.71	0.3	1
139	53	GLN	CA	C	55.809	0.3	1
140	53	GLN	H	H	9.134	0.02	1
141	53	GLN	N	N	116.73	0.3	1
142	54	HIS	CA	C	53.021	0.3	1
143	54	HIS	H	H	7.899	0.02	1
144	54	HIS	N	N	121.953	0.3	1
145	55	MET	CA	C	53.385	0.3	1
146	55	MET	H	H	7.384	0.02	1
147	55	MET	HA	H	4.411	0.02	1
148	55	MET	N	N	115.913	0.3	1
149	56	GLY	CA	C	44.418	0.3	1
150	56	GLY	H	H	7.319	0.02	1
151	56	GLY	HA2	H	3.688	0.02	1
152	56	GLY	N	N	107.419	0.3	1
153	57	PHE	CA	C	54.24	0.3	1
154	57	PHE	H	H	8.223	0.02	1
155	57	PHE	HA	H	4.126	0.02	1
156	57	PHE	N	N	114.088	0.3	1
157	58	LEU	CA	C	52.528	0.3	1
158	58	LEU	H	H	9.438	0.02	1
159	58	LEU	HA	H	5.159	0.02	1
160	58	LEU	N	N	121.131	0.3	1
161	59	CYS	CA	C	53.85	0.3	1
162	59	CYS	H	H	7.71	0.02	1
163	59	CYS	HA	H	4.159	0.02	1
164	59	CYS	N	N	115.179	0.3	1
165	60	ASN	CA	C	53.128	0.3	1
166	60	ASN	H	H	8.247	0.02	1

Atom shift assignment ID	Residue sequence code	Residue label	Atom name	Atom type	Chemical shift value (ppm)	Chemical shift value error (ppm)	Chemical shift ambiguity code
167	60	ASN	HA	H	3.776	0.02	1
168	60	ASN	N	N	119.48	0.3	1
169	66	ALA	CA	C	52.904	0.3	1
170	66	ALA	H	H	8.253	0.02	1
171	66	ALA	N	N	121.152	0.3	1
172	67	GLU	CA	C	60.087	0.3	1
173	67	GLU	H	H	12.31	0.02	1
174	67	GLU	HA	H	2.977	0.02	1
175	67	GLU	N	N	119.146	0.3	1
176	68	LEU	CA	C	55.48	0.3	1
177	68	LEU	H	H	6.873	0.02	1
178	68	LEU	HA	H	3.585	0.02	1
179	68	LEU	N	N	115.435	0.3	1
180	69	ARG	CA	C	58.07	0.3	1
181	69	ARG	H	H	7.589	0.02	1
182	69	ARG	N	N	118.09	0.3	1
183	70	PHE	CA	C	60.754	0.3	1
184	70	PHE	H	H	8.284	0.02	1
185	70	PHE	HA	H	3.561	0.02	1
186	70	PHE	N	N	120.74	0.3	1
187	71	LEU	CA	C	57.067	0.3	1
188	71	LEU	H	H	7.565	0.02	1
189	71	LEU	HA	H	3.558	0.02	1
190	71	LEU	N	N	119.178	0.3	1
191	72	ASP	CA	C	55.193	0.3	1
192	72	ASP	H	H	8.119	0.02	1
193	72	ASP	HA	H	4.043	0.02	1
194	72	ASP	N	N	118.19	0.3	1
195	73	LEU	CA	C	53.743	0.3	1
196	73	LEU	H	H	7.213	0.02	1
197	73	LEU	HA	H	4.217	0.02	1
198	73	LEU	N	N	117.908	0.3	1
199	74	VAL	CA	C	67.398	0.3	1
200	74	VAL	H	H	7.117	0.02	1
201	74	VAL	HA	H	3.223	0.02	1
202	74	VAL	N	N	120.217	0.3	1
203	75	PRO	CA	C	64.05	0.3	1
204	75	PRO	HA	H	3.975	0.02	1
205	76	SER	CA	C	59.417	0.3	1
206	76	SER	H	H	7.79	0.02	1
207	76	SER	HA	H	4.314	0.02	1
208	76	SER	N	N	112.727	0.3	1
209	77	LEU	CA	C	54.387	0.3	1
210	77	LEU	H	H	7.871	0.02	1
211	77	LEU	HA	H	3.874	0.02	1
212	77	LEU	N	N	120.741	0.3	1
213	78	GLN	CA	C	54.923	0.3	1
214	78	GLN	H	H	7.407	0.02	1
215	78	GLN	HA	H	3.808	0.02	1
216	78	GLN	N	N	111.22	0.3	1
217	79	LEU	CA	C	53.556	0.3	1
218	79	LEU	H	H	8.667	0.02	1
219	79	LEU	HA	H	3.719	0.02	1
220	79	LEU	N	N	118.405	0.3	1
221	80	ASP	CA	C	50.058	0.3	1
222	80	ASP	H	H	8.35	0.02	1

Atom shift assignment ID	Residue sequence code	Residue label	Atom name	Atom type	Chemical shift value (ppm)	Chemical shift value error (ppm)	Chemical shift ambiguity code
223	80	ASP	HA	H	4.883	0.02	1
224	80	ASP	N	N	125.951	0.3	1
225	82	ALA	CA	C	51.096	0.3	1
226	82	ALA	H	H	8.43	0.02	1
227	82	ALA	HA	H	4.418	0.02	1
228	82	ALA	N	N	121.109	0.3	1
229	83	GLN	CA	C	52.505	0.3	1
230	83	GLN	H	H	8.091	0.02	1
231	83	GLN	HA	H	3.77	0.02	1
232	83	GLN	N	N	119.862	0.3	1
233	84	ILE	CA	C	58.673	0.3	1
234	84	ILE	H	H	8.019	0.02	1
235	84	ILE	HA	H	4.273	0.02	1
236	84	ILE	N	N	118.119	0.3	1
237	85	TYR	CA	C	54.967	0.3	1
238	85	TYR	H	H	8.803	0.02	1
239	85	TYR	HA	H	5.361	0.02	1
240	85	TYR	N	N	122.093	0.3	1
241	86	ARG	CA	C	53.324	0.3	1
242	86	ARG	H	H	9.327	0.02	1
243	86	ARG	HA	H	5.243	0.02	1
244	86	ARG	N	N	121.357	0.3	1
245	87	VAL	CA	C	59.669	0.3	1
246	87	VAL	H	H	9.594	0.02	1
247	87	VAL	HA	H	4.924	0.02	1
248	87	VAL	N	N	131.988	0.3	1
249	88	THR	CA	C	60.903	0.3	1
250	88	THR	H	H	9.037	0.02	1
251	88	THR	HA	H	5.264	0.02	1
252	88	THR	N	N	124.077	0.3	1
253	89	TRP	CA	C	50.845	0.3	1
254	89	TRP	H	H	8.782	0.02	1
255	89	TRP	HA	H	3.934	0.02	1
256	89	TRP	N	N	124.495	0.3	1
257	90	PHE	CA	C	55.581	0.3	1
258	90	PHE	H	H	9.55	0.02	1
259	90	PHE	N	N	123.849	0.3	1
260	91	ILE	CA	C	57.163	0.3	1
261	91	ILE	H	H	9.617	0.02	1
262	91	ILE	HA	H	4.099	0.02	1
263	91	ILE	N	N	124.412	0.3	1
264	92	SER	CA	C	56.938	0.3	1
265	92	SER	H	H	9.027	0.02	1
266	92	SER	HA	H	5.816	0.02	1
267	92	SER	N	N	124.175	0.3	1
268	93	TRP	CA	C	53.565	0.3	1
269	93	TRP	H	H	6.608	0.02	1
270	93	TRP	N	N	116.806	0.3	1
271	94	SER	CA	C	57.036	0.3	1
272	94	SER	H	H	8.4	0.02	1
273	94	SER	HA	H	4.278	0.02	1
274	94	SER	N	N	114.978	0.3	1
275	97	PHE	CA	C	57.454	0.3	1
276	97	PHE	H	H	9.208	0.02	1
277	97	PHE	N	N	114.701	0.3	1
278	98	SER	CA	C	62.239	0.3	1

Atom shift assignment ID	Residue sequence code	Residue label	Atom name	Atom type	Chemical shift value (ppm)	Chemical shift value error (ppm)	Chemical shift ambiguity code
279	98	SER	H	H	10.833	0.02	1
280	98	SER	HA	H	3.713	0.02	1
281	98	SER	N	N	119.431	0.3	1
282	99	TRP	CA	C	58.178	0.3	1
283	99	TRP	H	H	7.904	0.02	1
284	99	TRP	HA	H	3.776	0.02	1
285	99	TRP	N	N	118.466	0.3	1
286	100	GLY	CA	C	44.939	0.3	1
287	100	GLY	H	H	7.377	0.02	1
288	100	GLY	HA2	H	3.693	0.02	1
289	100	GLY	N	N	105.865	0.3	1
290	101	CYS	CA	C	65.868	0.3	1
291	101	CYS	H	H	7.344	0.02	1
292	101	CYS	HA	H	3.859	0.02	1
293	101	CYS	N	N	120.612	0.3	1
294	102	ALA	CA	C	55.648	0.3	1
295	102	ALA	H	H	8.719	0.02	1
296	102	ALA	HA	H	3.865	0.02	1
297	102	ALA	N	N	119.725	0.3	1
298	103	GLY	CA	C	45.712	0.3	1
299	103	GLY	H	H	8.514	0.02	1
300	103	GLY	HA2	H	3.744	0.02	1
301	103	GLY	N	N	103.563	0.3	1
302	104	GLU	CA	C	58.268	0.3	1
303	104	GLU	H	H	7.899	0.02	1
304	104	GLU	HA	H	4.266	0.02	1
305	104	GLU	N	N	122.828	0.3	1
306	105	VAL	CA	C	65.964	0.3	1
307	105	VAL	H	H	8.894	0.02	1
308	105	VAL	HA	H	3.699	0.02	1
309	105	VAL	N	N	121.866	0.3	1
310	106	ARG	CA	C	59.471	0.3	1
311	106	ARG	H	H	8.567	0.02	1
312	106	ARG	HA	H	3.786	0.02	1
313	106	ARG	N	N	120.463	0.3	1
314	107	ALA	CA	C	54.029	0.3	1
315	107	ALA	H	H	7.764	0.02	1
316	107	ALA	HA	H	4.134	0.02	1
317	107	ALA	N	N	120.481	0.3	1
318	108	PHE	CA	C	59.933	0.3	1
319	108	PHE	H	H	8.093	0.02	1
320	108	PHE	HA	H	4.337	0.02	1
321	108	PHE	N	N	118.785	0.3	1
322	109	LEU	CA	C	56.159	0.3	1
323	109	LEU	H	H	8.595	0.02	1
324	109	LEU	HA	H	3.8	0.02	1
325	109	LEU	N	N	120.572	0.3	1
326	110	GLN	CA	C	55.701	0.3	1
327	110	GLN	H	H	8.082	0.02	1
328	110	GLN	HA	H	4.023	0.02	1
329	110	GLN	N	N	117.411	0.3	1
330	111	GLU	CA	C	54.17	0.3	1
331	111	GLU	H	H	7.336	0.02	1
332	111	GLU	HA	H	4.322	0.02	1
333	111	GLU	N	N	115.244	0.3	1
334	112	ASN	CA	C	50.899	0.3	1

Atom shift assignment ID	Residue sequence code	Residue label	Atom name	Atom type	Chemical shift value (ppm)	Chemical shift value error (ppm)	Chemical shift ambiguity code
335	112	ASN	H	H	7.561	0.02	1
336	112	ASN	HA	H	5.065	0.02	1
337	112	ASN	N	N	119.718	0.3	1
338	113	THR	CA	C	62.741	0.3	1
339	113	THR	H	H	8.01	0.02	1
340	113	THR	HA	H	4.317	0.02	1
341	113	THR	N	N	111.086	0.3	1
342	114	HIS	CA	C	55.367	0.3	1
343	114	HIS	H	H	8.436	0.02	1
344	114	HIS	HA	H	4.289	0.02	1
345	114	HIS	N	N	121.13	0.3	1
346	115	VAL	CA	C	60.602	0.3	1
347	115	VAL	H	H	7.273	0.02	1
348	115	VAL	HA	H	4.526	0.02	1
349	115	VAL	N	N	119.869	0.3	1
350	116	ARG	CA	C	52.095	0.3	1
351	116	ARG	H	H	8.729	0.02	1
352	116	ARG	HA	H	4.81	0.02	1
353	116	ARG	N	N	125.494	0.3	1
354	117	LEU	CA	C	52.214	0.3	1
355	117	LEU	H	H	8.663	0.02	1
356	117	LEU	HA	H	5.16	0.02	1
357	117	LEU	N	N	120.076	0.3	1
358	118	ARG	CA	C	53.466	0.3	1
359	118	ARG	H	H	8.974	0.02	1
360	118	ARG	HA	H	4.751	0.02	1
361	118	ARG	N	N	126.936	0.3	1
362	119	ILE	CA	C	59.937	0.3	1
363	119	ILE	H	H	8.946	0.02	1
364	119	ILE	HA	H	4.442	0.02	1
365	119	ILE	N	N	123.491	0.3	1
366	120	LYS	CA	C	53.021	0.3	1
367	120	LYS	H	H	9.12	0.02	1
368	120	LYS	N	N	128.843	0.3	1
369	121	ALA	CA	C	47.982	0.3	1
370	121	ALA	H	H	9.041	0.02	1
371	121	ALA	HA	H	5.67	0.02	1
372	121	ALA	N	N	124.831	0.3	1
373	122	ALA	CA	C	54.491	0.3	1
374	122	ALA	H	H	9.019	0.02	1
375	122	ALA	HA	H	5.973	0.02	1
376	122	ALA	N	N	125.972	0.3	1
377	123	ARG	CA	C	53.499	0.3	1
378	123	ARG	H	H	6.779	0.02	1
379	123	ARG	N	N	108.045	0.3	1
380	125	TYR	CA	C	55.356	0.3	1
381	125	TYR	H	H	8.785	0.02	1
382	125	TYR	HA	H	5.335	0.02	1
383	125	TYR	N	N	130.143	0.3	1
384	126	ASP	CA	C	53.234	0.3	1
385	126	ASP	H	H	7.98	0.02	1
386	126	ASP	HA	H	3.667	0.02	1
387	126	ASP	N	N	123.662	0.3	1
388	127	TYR	CA	C	58.15	0.3	1
389	127	TYR	H	H	6.226	0.02	1
390	127	TYR	HA	H	3.915	0.02	1

Atom shift assignment ID	Residue sequence code	Residue label	Atom name	Atom type	Chemical shift value (ppm)	Chemical shift value error (ppm)	Chemical shift ambiguity code
391	127	TYR	N	N	115.919	0.3	1
392	128	ASP	CA	C	49.091	0.3	1
393	128	ASP	H	H	6.897	0.02	1
394	128	ASP	N	N	117.348	0.3	1
395	129	PRO	CA	C	63.084	0.3	1
396	129	PRO	HA	H	4.103	0.02	1
397	130	LEU	CA	C	53.092	0.3	1
398	130	LEU	H	H	8.395	0.02	1
399	130	LEU	HA	H	4.698	0.02	1
400	130	LEU	N	N	120.168	0.3	1
401	131	TYR	CA	C	59.984	0.3	1
402	131	TYR	H	H	7.956	0.02	1
403	131	TYR	HA	H	4.061	0.02	1
404	131	TYR	N	N	117.484	0.3	1
405	132	LYS	CA	C	59.01	0.3	1
406	132	LYS	H	H	7.461	0.02	1
407	132	LYS	N	N	121.876	0.3	1
408	133	GLU	CA	C	59.291	0.3	1
409	133	GLU	H	H	8.716	0.02	1
410	133	GLU	N	N	121.833	0.3	1
411	134	ALA	CA	C	54.173	0.3	1
412	134	ALA	H	H	7.723	0.02	1
413	134	ALA	HA	H	3.846	0.02	1
414	134	ALA	N	N	119.672	0.3	1
415	135	LEU	CA	C	57.172	0.3	1
416	135	LEU	H	H	7.211	0.02	1
417	135	LEU	N	N	116.268	0.3	1
418	136	GLN	CA	C	58.139	0.3	1
419	136	GLN	H	H	8.349	0.02	1
420	136	GLN	HA	H	3.719	0.02	1
421	136	GLN	N	N	117.393	0.3	1
422	137	MET	CA	C	58.155	0.3	1
423	137	MET	H	H	8.187	0.02	1
424	137	MET	HA	H	3.917	0.02	1
425	137	MET	N	N	118.442	0.3	1
426	138	LEU	CA	C	56.955	0.3	1
427	138	LEU	H	H	7.928	0.02	1
428	138	LEU	HA	H	4.335	0.02	1
429	138	LEU	N	N	118.846	0.3	1
430	139	ARG	CA	C	57.716	0.3	1
431	139	ARG	H	H	8.004	0.02	1
432	139	ARG	HA	H	4.426	0.02	1
433	139	ARG	N	N	121.137	0.3	1
434	140	ASP	CA	C	56.076	0.3	1
435	140	ASP	H	H	8.816	0.02	1
436	140	ASP	HA	H	4.429	0.02	1
437	140	ASP	N	N	122.344	0.3	1
438	141	ALA	CA	C	50.807	0.3	1
439	141	ALA	H	H	7.708	0.02	1
440	141	ALA	HA	H	4.418	0.02	1
441	141	ALA	N	N	119.974	0.3	1
442	142	GLY	CA	C	43.433	0.3	1
443	142	GLY	H	H	7.926	0.02	1
444	142	GLY	HA2	H	3.722	0.02	1
445	142	GLY	N	N	104.888	0.3	1
446	143	ALA	CA	C	50.331	0.3	1

Atom shift assignment ID	Residue sequence code	Residue label	Atom name	Atom type	Chemical shift value (ppm)	Chemical shift value error (ppm)	Chemical shift ambiguity code
447	143	ALA	H	H	8.031	0.02	1
448	143	ALA	HA	H	4.689	0.02	1
449	143	ALA	N	N	124.036	0.3	1
450	144	GLN	CA	C	53.671	0.3	1
451	144	GLN	H	H	8.1	0.02	1
452	144	GLN	HA	H	4.439	0.02	1
453	144	GLN	N	N	121.862	0.3	1
454	145	VAL	CA	C	59.951	0.3	1
455	145	VAL	H	H	9.01	0.02	1
456	145	VAL	HA	H	4.752	0.02	1
457	145	VAL	N	N	129.096	0.3	1
458	146	SER	CA	C	56.265	0.3	1
459	146	SER	H	H	9.235	0.02	1
460	146	SER	HA	H	5.179	0.02	1
461	146	SER	N	N	120.374	0.3	1
462	149	THR	CA	C	58.56	0.3	1
463	149	THR	H	H	8.473	0.02	1
464	149	THR	N	N	117.515	0.3	1
465	150	TYR	CA	C	62.747	0.3	1
466	150	TYR	H	H	9.687	0.02	1
467	150	TYR	N	N	121.557	0.3	1
468	151	ASP	CA	C	56.036	0.3	1
469	151	ASP	H	H	8.654	0.02	1
470	151	ASP	HA	H	5.435	0.02	1
471	151	ASP	N	N	115.195	0.3	1
472	155	TYR	CA	C	60.273	0.3	1
473	155	TYR	H	H	8.013	0.02	1
474	155	TYR	N	N	121.106	0.3	1
475	156	CYS	CA	C	63.256	0.3	1
476	156	CYS	H	H	7.973	0.02	1
477	156	CYS	HA	H	3.789	0.02	1
478	156	CYS	N	N	118.677	0.3	1
479	157	TRP	CA	C	58.296	0.3	1
480	157	TRP	H	H	7.911	0.02	1
481	157	TRP	HA	H	3.764	0.02	1
482	157	TRP	N	N	122.741	0.3	1
483	158	ASP	CA	C	54.888	0.3	1
484	158	ASP	H	H	8.038	0.02	1
485	158	ASP	HA	H	5.995	0.02	1
486	158	ASP	N	N	113.827	0.3	1
487	159	THR	CA	C	64.329	0.3	1
488	159	THR	H	H	7.443	0.02	1
489	159	THR	HA	H	5.439	0.02	1
490	159	THR	N	N	115.142	0.3	1
491	160	PHE	CA	C	57.29	0.3	1
492	160	PHE	H	H	8.215	0.02	1
493	160	PHE	HA	H	4.475	0.02	1
494	160	PHE	N	N	114.292	0.3	1
495	161	VAL	CA	C	59.495	0.3	1
496	161	VAL	H	H	6.519	0.02	1
497	161	VAL	HA	H	4.025	0.02	1
498	161	VAL	N	N	117.838	0.3	1
499	162	TYR	CA	C	54.861	0.3	1
500	162	TYR	H	H	8.769	0.02	1
501	162	TYR	HA	H	4.753	0.02	1
502	162	TYR	N	N	128.373	0.3	1

Atom shift assignment ID	Residue sequence code	Residue label	Atom name	Atom type	Chemical shift value (ppm)	Chemical shift value error (ppm)	Chemical shift ambiguity code
503	163	ARG	CA	C	56.904	0.3	1
504	163	ARG	H	H	8.428	0.02	1
505	163	ARG	HA	H	4.1	0.02	1
506	163	ARG	N	N	127.444	0.3	1
507	165	GLY	CA	C	44.824	0.3	1
508	165	GLY	H	H	11.556	0.02	1
509	165	GLY	HA2	H	3.702	0.02	1
510	165	GLY	N	N	111.911	0.3	1
511	166	CYS	CA	C	54.18	0.3	1
512	166	CYS	H	H	7.916	0.02	1
513	166	CYS	HA	H	4.716	0.02	1
514	166	CYS	N	N	123.647	0.3	1
515	168	PHE	CA	C	56.234	0.3	1
516	168	PHE	H	H	9.134	0.02	1
517	168	PHE	HA	H	6.488	0.02	1
518	168	PHE	N	N	121.572	0.3	1
519	169	GLN	CA	C	49.506	0.3	1
520	169	GLN	H	H	7.184	0.02	1
521	169	GLN	HA	H	4.317	0.02	1
522	169	GLN	N	N	128.787	0.3	1
523	171	TRP	CA	C	53.634	0.3	1
524	171	TRP	H	H	5.542	0.02	1
525	171	TRP	HA	H	4.633	0.02	1
526	171	TRP	N	N	117.476	0.3	1
527	172	ASP	CA	C	52.154	0.3	1
528	172	ASP	H	H	9.245	0.02	1
529	172	ASP	HA	H	5.079	0.02	1
530	172	ASP	N	N	125.428	0.3	1
531	174	LEU	CA	C	59.641	0.3	1
532	174	LEU	H	H	6.878	0.02	1
533	174	LEU	HA	H	4.346	0.02	1
534	174	LEU	N	N	119.89	0.3	1
535	175	GLU	CA	C	59.19	0.3	1
536	175	GLU	H	H	9.377	0.02	1
537	175	GLU	HA	H	3.723	0.02	1
538	175	GLU	N	N	119.587	0.3	1
539	176	GLU	CA	C	59.094	0.3	1
540	176	GLU	H	H	8.28	0.02	1
541	176	GLU	HA	H	5.195	0.02	1
542	176	GLU	N	N	122.831	0.3	1
543	177	HIS	CA	C	57.678	0.3	1
544	177	HIS	H	H	8.812	0.02	1
545	177	HIS	HA	H	4.212	0.02	1
546	177	HIS	N	N	120.669	0.3	1
547	178	SER	CA	C	61.272	0.3	1
548	178	SER	H	H	8.476	0.02	1
549	178	SER	N	N	115.134	0.3	1
550	179	GLN	CA	C	58.067	0.3	1
551	179	GLN	H	H	8.53	0.02	1
552	179	GLN	N	N	121.487	0.3	1
553	180	ALA	CA	C	54.124	0.3	1
554	180	ALA	H	H	8.293	0.02	1
555	180	ALA	HA	H	4.284	0.02	1
556	180	ALA	N	N	124.243	0.3	1
557	182	SER	CA	C	61.46	0.3	1
558	182	SER	H	H	8.905	0.02	1

Atom shift assignment ID	Residue sequence code	Residue label	Atom name	Atom type	Chemical shift value (ppm)	Chemical shift value error (ppm)	Chemical shift ambiguity code
559	182	SER	HA	H	4.039	0.02	1
560	182	SER	N	N	116.856	0.3	1
561	183	GLY	CA	C	46.182	0.3	1
562	183	GLY	H	H	8.18	0.02	1
563	183	GLY	HA2	H	4.147	0.02	1
564	183	GLY	N	N	109.656	0.3	1
565	184	ARG	CA	C	58.273	0.3	1
566	184	ARG	H	H	7.738	0.02	1
567	184	ARG	HA	H	4.237	0.02	1
568	184	ARG	N	N	122.994	0.3	1
569	185	LEU	CA	C	56.989	0.3	1
570	185	LEU	H	H	8.736	0.02	1
571	185	LEU	HA	H	4.238	0.02	1
572	185	LEU	N	N	119.765	0.3	1
573	186	ARG	CA	C	58.676	0.3	1
574	186	ARG	H	H	8.331	0.02	1
575	186	ARG	HA	H	4.135	0.02	1
576	186	ARG	N	N	118.311	0.3	1
577	187	ALA	CA	C	53.493	0.3	1
578	187	ALA	H	H	7.355	0.02	1
579	187	ALA	HA	H	4.29	0.02	1
580	187	ALA	N	N	119.351	0.3	1
581	188	ILE	CA	C	63.627	0.3	1
582	188	ILE	H	H	7.87	0.02	1
583	188	ILE	HA	H	4.421	0.02	1
584	188	ILE	N	N	119.106	0.3	1
585	189	LEU	CA	C	54.295	0.3	1
586	189	LEU	H	H	8.164	0.02	1
587	189	LEU	HA	H	4.108	0.02	1
588	189	LEU	N	N	115.663	0.3	1

Table 8.3 Kinetic parameters of A3 variants analysed using linear and non-linear regression methods (I acknowledge Prof. Geoffrey B. Jameson for conducting non-linear regression analysis)

Enzyme	Oligonucleotide	Parameters	Linear analysis	Non-linear
$A3B_{CTD}$-AL1	T**C**A-oligo	K_m (μM)	197 ± 30	146 ± 34
		V_{max} (μM/s)	0.014 ± 0.002	0.013 ± 0.001
$A3B_{CTD}$-DM	T**C**A-oligo	K_m (μM)	317 ± 55	438 ± 105
		V_{max} (μM/s)	0.017 ± 0.003	0.021 ± 0.003
A3A-mimic	CC**C**-oligo	K_m (μM)	440 ± 80	629 ± 211
($A3B_{CTD}$-AL1)		V_{max} (μM/s)	0.019 ± 0.003	0.025 ± 0.005
$A3B_{CTD}$-DM	**C**CC-oligo	K_m (μM)	590 ± 90	646 ± 120
		V_{max} (μM/s)	0.017 ± 0.003	0.019 ± 0.002
A3A-mimic	dZC**C**-oligo	K_m (μM)	520 ± 90	784 ± 210
($A3B_{CTD}$-AL1)		V_{max} (μM/s)	0.022 ± 0.005	0.030 ± 0.005
$A3B_{CTD}$-DM	**C**CdZ-oligo	K_m (μM)	1370 ± 270	1072 ± 296
		V_{max} (μM/s)	0.020 ± 0.004	0.018 ± 0.002
$A3G_{CTD}$-Wt	dZC**C**-oligo	K_m (μM)	15805 ± 1305	1174 ± 415
		V_{max} (μM/s)	0.088 ± 0.087	0.082 ± 0.014

Table 8.4 Inhibition constants and parameters of A3 variants analysed using linear and non-linear regression methods (I acknowledge Prof. Geoffrey B. Jameson for conducting non-linear regression analysis)

Enzyme	Oligonucleotide	Parameters	Linear analysis	Non-linear
$A3B_{CTD}$-AL1	**dZ**-oligo	K_i (μM)	7.5 ± 1.7	8.0 ± 1.5
		V_{max} (μM/s)	0.0095 ± 0.0009	0.0143 ± 0.0010
$A3B_{CTD}$-DM	**dZ**-oligo	K_i (μM)	11.4 ± 2.6	8.6 ± 1.2
		V_{max} (μM/s)	0.0063 ± 0.0004	0.0133 ± 0.0005
$A3B_{CTD}$-AL1	**5FdZ**-oligo	K_i (μM)	2.1 ± 0.8	4.6 ± 1.1
		V_{max} (μM/s)	0.016 ± 0.007	0.015 ± 0.0016
A3A-mimic	CC**dZ**-oligo	K_i (μM)	12.3 ± 1.8	11.7 ± 2.9
($A3B_{CTD}$-AL1)		V_{max} (μM/s)	0.0057 ± 0.0034	0.0088 ± 0.0006
$A3B_{CTD}$-DM	**dZ**CC-oligo	K_i (μM)	18.5 ± 5.5	22.2 ± 2.3
		V_{max} (μM/s)	0.00823 ± 0.0010	0.0145 ± 0.0006

Table 8.5 ^{1}H-NMR signal location of H-5 doublets of dC and dU in ssDNAs and their J-coupling

Oligonucleotide	^{1}H Signal	Chemical shifts (ppm)	J-coupling (Hz)
A9C-oligo	H-5 dC	5.84	7.50
A9U-oligo	H-5 dU	5.665	7.65
A9C-oligo	H-4 dT	7.41	7.72
9merC--oligo	H-5 dC	5.90	7.55
9merU-oligo	H-5 dU	5.72	7.63
CCU-oligo	H-5 dU	5.673	8.49
CUC-oligo	H-5 dU	5.748	8.43
UCC-oligo	H-5 dU	5.775	8.5

Table 8.6 Interface residue contacts of $A3B_{CTD}$-QM-ΔL3 (5CQI, PDB) interface A (Δ*G* = -12.65 kcal/mol) (PRISM [9])

Number	$A3B_{CTD}$_A	Residue contact	$A3B_{CTD}$ _B
1	pdb1_A_TYR_350	←→	pdb2_B_ASP_194
2	pdb1_A_TYR_215	←→	pdb2_B_TYR_191
3	pdb1_A_SER_264	←→	pdb2_B_SER_264
4	pdb1_A_SER_264	←→	pdb2_B_VAL_262
5	pdb1_A_PRO_263	←→	pdb2_B_PRO_263
6	pdb1_A_SER_264	←→	pdb2_B_ASP_260
7	pdb1_A_SER_264	←→	pdb2_B_LEU_261
8	pdb1_A_GLN_233	←→	pdb2_B_GLU_241
9	pdb1_A_ASP_232	←→	pdb2_B_GLU_241
10	pdb1_A_LEU_261	←→	pdb2_B_SER_264
11	pdb1_A_TYR_191	←→	pdb2_B_PHE_348
12	pdb1_A_ASP_260	←→	pdb2_B_GLN_266
13	pdb1_A_GLN_266	←→	pdb2_B_ASP_260
14	pdb1_A_MET_193	←→	pdb2_B_TYR_191
15	pdb1_A_SER_264	←→	pdb2_B_PRO_263
16	pdb1_A_ASP_194	←→	pdb2_B_TYR_350
17	pdb1_A_GLN_233	←→	pdb2_B_ARG_257
18	pdb1_A_LEU_192	←→	pdb2_B_TYR_191
19	pdb1_A_TYR_191	←→	pdb2_B_TYR_191
20	pdb1_A_TYR_191	←→	pdb2_B_MET_193
21	pdb1_A_TYR_191	←→	pdb2_B_LEU_192
22	pdb1_A_TYR_191	←→	pdb2_B_PHE_237

Table 8.7 Interface residue contacts of A3B$_{CTD}$-QM-ΔL3 (5CQI, PDB) interface B (Δ*G* = -2.94 kcal/mol) (PRISM [9])

Number	A3B$_{CTD}$_A	Residue contact	A3B$_{CTD}$ _B
1	pdb1_A_GLY_251	←→	pdb2_B_TYR_191
2	pdb1_A_GLN_213	←→	pdb2_B_ASP_194
3	pdb1_A_HIS_253	←→	pdb2_B_TYR_350
4	pdb1_A_ARG_212	←→	pdb2_B_PRO_195
5	pdb1_A_ARG_212	←→	pdb2_B_ASP_194
6	pdb1_A_ARG_212	←→	pdb2_B_THR_197
7	pdb1_A_ARG_212	←→	pdb2_B_ASP_196
8	pdb1_A_SER_250	←→	pdb2_B_MET_193
9	pdb1_A_GLY_251	←→	pdb2_B_LEU_192
10	pdb1_A_GLU_241	←→	pdb2_B_PHE_237
11	pdb1_A_TYR_191	←→	pdb2_B_SER_250
12	pdb1_A_SER_250	←→	pdb2_B_PHE_237
13	pdb1_A_GLY_251	←→	pdb2_B_TYR_350
14	pdb1_A_TRP_287	←→	pdb2_B_TYR_350
15	pdb1_A_TRP_287	←→	pdb2_B_GLN_352

Table 8.8 Interface residue contacts of A3B$_{CTD}$-AL1 interface model 1 (Δ*G* = -38.13 kcal/mol) (PRISM [9])

Number	A3B$_{CTD}$-AL1_A	Residue contact	A3B$_{CTD}$-AL1_B
1	pdb1_A_ASP_260	←→	pdb2_B_TYR_350
2	pdb1_A_CYS_239	←→	pdb2_B_CYS_239
3	pdb1_A_LYS_213	←→	pdb2_B_LYS_213
4	pdb1_A_TYR_191	←→	pdb2_B_ARG_257
5	pdb1_A_GLU_241	←→	pdb2_B_TYR_215
6	pdb1_A_ASP_196	←→	pdb2_B_SER_250
7	pdb1_A_ARG_257	←→	pdb2_B_THR_197
8	pdb1_A_THR_197	←→	pdb2_B_GLU_241
9	pdb1_A_ASP_260	←→	pdb2_B_ASP_194
10	pdb1_A_ARG_257	←→	pdb2_B_MET_193
11	pdb1_A_TYR_215	←→	pdb2_B_SER_250
12	pdb1_A_MET_193	←→	pdb2_B_ARG_257
13	pdb1_A_TYR_191	←→	pdb2_B_LEU_261
14	pdb1_A_PHE_237	←→	pdb2_B_GLU_241
15	pdb1_A_MET_235	←→	pdb2_B_TYR_191
16	pdb1_A_TYR_218	←→	pdb2_B_TYR_191
17	pdb1_A_LEU_261	←→	pdb2_B_TYR_191
18	pdb1_A_LEU_265	←→	pdb2_B_TYR_191
19	pdb1_A_ARG_257	←→	pdb2_B_PHE_237
20	pdb1_A_TYR_215	←→	pdb2_B_GLU_241
21	pdb1_A_MET_193	←→	pdb2_B_GLU_241
22	pdb1_A_THR_197	←→	pdb2_B_SER_250
23	pdb1_A_GLU_241	←→	pdb2_B_THR_197
24	pdb1_A_PHE_237	←→	pdb2_B_ARG_257

Number	$A3B_{CTD}$-AL1_A	Residue contact	$A3B_{CTD}$-AL1_B
25	pdb1_A_ASP_194	←→	pdb2_B_ARG_257
26	pdb1_A_SER_250	←→	pdb2_B_LYS_213
27	pdb1_A_ASP_194	←→	pdb2_B_ARG_252
28	pdb1_A_ARG_257	←→	pdb2_B_ASP_194
29	pdb1_A_SER_264	←→	pdb2_B_ARG_190
30	pdb1_A_SER_264	←→	pdb2_B_TYR_191
31	pdb1_A_SER_264	←→	pdb2_B_LEU_192
32	pdb1_A_TYR_191	←→	pdb2_B_LEU_238
33	pdb1_A_PRO_263	←→	pdb2_B_TYR_350
34	pdb1_A_ASN_201	←→	pdb2_B_SER_250
35	pdb1_A_ARG_252	←→	pdb2_B_ASP_196
36	pdb1_A_HIS_234	←→	pdb2_B_TYR_191

Table 8.9 Interface residue contacts of $A3B_{CTD}$-AL1 interface model 2 (ΔG = -18.6 kcal/mol) (PRISM [9])

Number	$A3B_{CTD}$-AL1_A	Residue contact	$A3B_{CTD}$-AL1_B
1	pdb1_A_GLY_291	←→	pdb2_B_GLY_288
2	pdb1_A_GLU_299	←→	pdb2_B_ARG_252
3	pdb1_A_GLU_299	←→	pdb2_B_GLY_251
4	pdb1_A_GLU_299	←→	pdb2_B_SER_250
5	pdb1_A_GLY_291	←→	pdb2_B_SER_286
6	pdb1_A_GLY_291	←→	pdb2_B_TRP_287
7	pdb1_A_TRP_287	←→	pdb2_B_ALA_295
8	pdb1_A_ALA_295	←→	pdb2_B_TRP_287
9	pdb1_A_MET_325	←→	pdb2_B_SER_286
10	pdb1_A_ASP_260	←→	pdb2_B_ARG_252
11	pdb1_A_GLU_292	←→	pdb2_B_LEU_256
12	pdb1_A_GLU_292	←→	pdb2_B_ARG_252
13	pdb1_A_PHE_296	←→	pdb2_B_SER_250
14	pdb1_A_PHE_296	←→	pdb2_B_GLY_251
15	pdb1_A_ALA_295	←→	pdb2_B_HIS_253
16	pdb1_A_SER_250	←→	pdb2_B_GLU_299
17	pdb1_A_PRO_263	←→	pdb2_B_SER_250
18	pdb1_A_PHE_285	←→	pdb2_B_PHE_285
19	pdb1_A_GLU_299	←→	pdb2_B_GLU_241
20	pdb1_A_GLU_299	←→	pdb2_B_ASN_240
21	pdb1_A_ASN_300	←→	pdb2_B_SER_250
22	pdb1_A_ASN_300	←→	pdb2_B_GLY_251
23	pdb1_A_GLU_292	←→	pdb2_B_TRP_287
24	pdb1_A_ASP_260	←→	pdb2_B_ASP_260
25	pdb1_A_LEU_256	←→	pdb2_B_GLU_292
26	pdb1_A_LEU_259	←→	pdb2_B_ARG_252
27	pdb1_A_GLY_251	←→	pdb2_B_GLU_299

Table 8.10 Model 1 dimer interface assessment using PISA [132] (Δ*G* = -2.00 kcal/mol, interface area = 1034.4 Å^2)

Number	$A3B_{CTD}$-AL1_A	Residue contact	$A3B_{CTD}$-AL1_B
1	ARG 257	← H-bond →	THR 197
2	THR 197	← H-bond →	GLU 241
3	SER 264	← H-bond →	ARG 190
4	GLU 241	← H-bond →	CYS 239
5	ASP 194	← H-bond →	ARG 252
6	ASP 196	← H-bond →	ARG 252
7	LEU 192	← H-bond →	ARG 257
8	ARG 252	← Salt-bridge →	ASP 196
9	ASP 194	← Salt-bridge →	ARG 252
10	ASP 196	← Salt-bridge →	ARG 252

Table 8.11 Model 2 dimer interface assessment PISA [132] (Δ*G* = -4.00 kcal/mol, interface area = 725.8 Å^2)

Number	$A3B_{CTD}$-AL1_A	Residue contact	$A3B_{CTD}$-AL1_B
1	GLU 292	← H-bond →	ARG 252
2	GLU 299	← H-bond →	SER 250
3	ASP 260	← Salt-bridge →	ARG 252
4	GLU 292	← Salt-bridge →	ARG 252

Table 8.12 SAXS *ab intio* modelling parameters of the A3Gfl_Wt along with CCdZ-40mer oligo

Modelling parameters	**A3G_species 2**	**A3G_species 3**	**A3G_species 4**
Ab intio restoration			
DAMMIF (default parameters, 10 calculations)			
q range (Å^{-1}) for fitting	0.0139 - 0.3052	0.0183 - 0.3052	0.0183 - 0.3052
Symmetry	P4	P2	P1
NSD (standard deviation)	0.578 (0.066)	0.937 (0.126)	0.559 (0.049)
Resolution (from SASRES) (Å)	36 ± 3	47 ± 4	35 ± 3

Table 8.13 SAXS fitting and modelling parameters of the A3Gfl_Wt along with CCdZ-40mer oligo to reported A3G dimer crystal structures

Modelling parameters	**A3G_species 3**	**A3G_species 4**
Structure modelling		
PDB, 6p40		
FoXS, Chi^2	0.50	4.93
CRYSOL, Chi^2	0.34	2.39
PDB, 6p3x		
FoXS, Chi^2	1.22	5.87
CRYSOL, Chi^2	0.71	2.86
PDB, 6p3y		
FoXS, Chi^2	1.27	5.95
CRYSOL, Chi^2	0.72	2.89
PDB, 6p3z		
FoXS, Chi^2	1.28	5.95
CRYSOL, Chi^2	0.73	2.89

9. References

1. Alexandrov, L. B.; Nik-Zainal, S.; Wedge, D. C.; Aparicio, S. A. J. R.; Behjati, S.; Biankin, A. V.; Bignell, G. R.; Bolli, N.; Borg, A.; Borresen-Dale, A.-L.; Boyault, S.; Burkhardt, B.; Butler, A. P.; Caldas, C.; Davies, H. R.; Desmedt, C., et al., Signatures of mutational processes in human cancer. *Nature* **2013,** *500* (7463), 415-421.
2. Anant, S.; MacGinnitie, A. J.; Davidson, N. O., apobec-1, the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme, is a novel RNA-binding protein. *Journal of Biological Chemistry* **1995,** *270* (24), 14762-14767.
3. Aoufouchi, S.; Faili, A.; Zober, C.; d'Orlando, O.; Weller, S.; Weill, J.-C.; Reynaud, C.-A., Proteasomal degradation restricts the nuclear lifespan of AID. *Journal of Experimental Medicine* **2008,** *205* (6), 1357-1368.
4. Arakawa, H.; Hauschild J.; Buerstedde, J.-M., Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. *Science* **2004,** *295* (1301), 1301-1305.
5. Azatian, S. B.; Kaur, N.; Latham, M. P., Increasing the buffering capacity of minimal media leads to higher protein yield. *Journal of biomolecular NMR* **2019**, 1-7.
6. Barchi Jr, J. J.; Haces, A.; Marquez, V. E.; McCormack, J. J., Inhibition of cytidine deaminase by derivatives of 1-(β-d-ribofuranosyl)-dihydropyrimidin-2-one (zebularine). *Nucleosides and Nucleotides* **1992,** *11* (10), 1781-1793.
7. Barnor, J. S.; Miyano-Kurosaki, N.; Yamaguchi, K.; Sakamoto, A.; Ishikawa, K.; Inagaki, Y.; Yamamoto, N.; Osei-Kwasi, M.; Ofori-Adjei, D.; Takaku, H., Intracellular expression of antisense RNA transcripts complementary to the human immunodeficiency virus type-1 vif gene inhibits viral replication in infected T-lymphoblastoid cells. *Biochemical and Biophysical Research Communications* **2004,** *320* (2), 544-550.
8. Barzak, F. M.; Harjes, S.; Kvach, M. V.; Kurup, H. M.; Jameson, G. B.; Filichev, V. V.; Harjes, E., Selective inhibition of APOBEC3 enzymes by single-stranded DNAs containing 2′-deoxyzebularine. *Organic and Biomolecular Chemistry* **2019,** *17* (43), 9435-9441.
9. Baspinar, A.; Cukuroglu, E.; Nussinov, R.; Keskin, O.; Gursoy, A., PRISM: a web server and repository for prediction of protein–protein interactions and modeling their 3D complexes. *Nucleic acids research* **2014,** *42* (W1), W285-W289.
10. Baumert, T. F.; Rösler, C.; Malim, M. H.; von Weizsäcker, F., Hepatitis B virus DNA is subject to extensive editing by the human deaminase APOBEC3C. . *Hepatology* **2007,** *46* (3), 682–689.
11. Beale, R. C. L.; Petersen-Mahrt, S. K.; Watt, I. N.; Harris, R. S.; Rada, C.; Neuberger, M. S., Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. *Journal of Molecular Biology* **2004,** *337* (3), 585-596.
12. Bennett R. P.; Presnyak V.; Wedekind J. E.; Smith H. C., Nuclear exclusion of the HIV-1 host defense factor APOBEC3G requires a novel cytoplasmic retention signal and is not dependent on RNA binding. *Journal of Biological Chemistry* **2007,** *283* (12), 7320-7327.
13. Bennett, R. P.; Salter, J. D.; Liu, X.; Wedekind, J. E.; Smith, H. C., APOBEC3G subunits self-associate via the C-terminal deaminase domain. *Journal of Biological Chemistry* **2008,** *283* (48), 33329-33336.
14. Betts, L.; Xiang, S.; Short, S. A.; Wolfenden, R.; Carter, C. W., Cytidine deaminase. The 2·3 Å crystal structure of an enzyme: transition-state analog complex. *Journal of Molecular Biology* **1994,** *235* (2), 635-656.

15. Bhattacharya, S.; Navaratnam, N.; Morrison, J. R.; Scott, J.; Taylor, W. R., Cytosine nucleoside/nucleotide deaminases and apolipoprotein B mRNA editing. *Trends in biochemical sciences* **1994,** *19* (3), 105.
16. Bishop, K. N.; Holmes, R. K.; Malim, M. H., Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. *Journal of virology* **2006,** *80* (17), 8450-8458.
17. Bishop, K. N.; Holmes, R. K.; Sheehy, A. M.; Davidson, N. O.; Cho, S.-J.; Malim, M. H., Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Current Biology* **2004,** *14* (15), 1392-1396.
18. Bishop, K. N.; Verma, M.; Kim, E.-Y.; Wolinsky, S. M.; Malim, M. H., APOBEC3G inhibits elongation of HIV-1 reverse transcripts. *PLoS pathogens* **2008,** *4* (12).
19. Blanc, V.; Davidson, N. O., APOBEC-1-mediated RNA editing. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **2010,** *2* (5), 594-602.
20. Bogerd, H. P.; Cullen, B. R., Single-stranded RNA facilitates nucleocapsid: APOBEC3G complex formation. *Rna* **2008,** *14* (6), 1228-1236.
21. Bogerd, H. P.; Wiegand, H. L.; Doehle, B. P.; Lueders, K. K.; Cullen, B. R., APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. *Nucleic acids research* **2006,** *34* (1), 89-95.
22. Bohn, M.-F.; Shandilya, S. M.; Albin, J. S.; Kouno, T.; Anderson, B. D.; McDougle, R. M.; Carpenter, M. A.; Rathore, A.; Evans, L.; Davis, A. N., Crystal structure of the DNA cytosine deaminase APOBEC3F: the catalytically active and HIV-1 Vif-binding domain. *Structure* **2013,** *21* (6), 1042-1050.
23. Bohn, M.-F.; Shandilya, S. M.; Silvas, T. V.; Nalivaika, E. A.; Kouno, T.; Kelch, B. A.; Ryder, S. P.; Kurt-Yilmaz, N.; Somasundaran, M.; Schiffer, C. A., The ssDNA mutator APOBEC3A is regulated by cooperative dimerization. *Structure* **2015,** *23* (5), 903-911.
24. Bonvin, M.; Greeve, J., Effects of point mutations in the cytidine deaminase domains of APOBEC3B on replication and hypermutation of hepatitis B virus in vitro. *Journal of General Virology* **2007,** *88* (12), 3270-3274.
25. Bradford M., A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. . *Analytical Biochemistry* **1976,** *72*, 248-254.
26. Bransteitter, R.; Pham, P.; Scharff, M. D.; Goodman, M. F., Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proceedings of the National Academy of Sciences USA* **2003,** *100* (7), 4102-4107.
27. Brar, S. S.; Sacho, E. J.; Tessmer, I.; Croteau, D. L.; Erie, D. A.; Diaz, M., Activation-induced deaminase, AID, is catalytically active as a monomer on single-stranded DNA. *DNA repair* **2008,** *7* (1), 77-87.
28. Brar, S. S.; Watson, M.; Diaz, M., Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks. *Journal of Biological Chemistry* **2004,** *279* (25), 26395-26401.
29. Bravo, I. G.; Félez-Sánchez, M., Papillomaviruses Viral evolution, cancer and evolutionary medicine. *Evolution, medicine, and public health* **2015,** *2015* (1), 32-51.
30. Breasted, J. H., The Edwin Smith Surgical Papyrus. In *The Edwin Smith Surgical Papyrus*, The University of Chicago Press: In facsimile and hieroglyphic transliteration with translation and commentary in two volumes, 1931; Vol. 3.
31. Brookes, E.; Pérez, J.; Cardinali, B.; Profumo, A.; Vachette, P.; Rocco, M., Fibrinogen species as resolved by HPLC-SAXS data processing within the UltraScan

Solution Modeler (US-SOMO) enhanced SAS module. *Journal of applied crystallography* **2013,** *46* (6), 1823-1833.
32. Brookes, E.; Rocco, M., Recent advances in the UltraScan SOlution MOdeller (US-SOMO) hydrodynamic and small-angle scattering data analysis and simulation suite. *European Biophysics Journal* **2018,** *47* (7), 855-864.
33. Brookes, E.; Vachette, P.; Rocco, M.; Pérez, J., US-SOMO HPLC-SAXS module: dealing with capillary fouling and extraction of pure component patterns from poorly resolved SEC-SAXS data. *Journal of applied crystallography* **2016,** *49* (5), 1827-1841.
34. Burns, M. B.; Lackey, L.; Carpenter, M. A.; Rathore, A.; Land, A. M.; Leonard, B.; Refsland, E. W.; Kotandeniya, D.; Tretyakova, N.; Nikas, J. B., APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **2013,** *494* (7437), 366-370.
35. Burns, M. B.; Temiz, N. A.; Harris, R. S., Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nature Genetics* **2013b,** *45* (9), 977-983.
36. Byeon I-J.L.; Ahn J.; Mitra M.; Byeon C-H.; Herc K.; Hritz1 J.; Charlton L.M.; Levin J.G.; Gronenborn A.M., NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity. *Nature communication* **2013,** *4*.
37. Byeon, I.-J. L.; Byeon, C.-H.; Wu, T.; Mitra, M.; Singer, D.; Levin, J. G.; Gronenborn, A. M., Nuclear magnetic resonance structure of the APOBEC3B catalytic domain: structural basis for substrate binding and DNA deaminase activity. *Biochemistry* **2016,** *55* (21), 2944-2959.
38. Bzhalava, D.; Eklund, C.; Dillner, J., International standardization and classification of human papillomavirus types. *Virology* **2015,** *476*, 341-344.
39. Cancer Genome Atlas Research Network, Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **2014,** *507* (7492), 315-322.
40. Carpenter, M. A.; Li, M.; Rathore, A.; Lackey, L.; Law, E. K.; Land, A. M.; Leonard, B.; Shandilya, S. M.; Bohn, M.-F.; Schiffer, C. A., Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A. *Journal of Biological Chemistry* **2012,** *287* (41), 34801-34808.
41. Carpenter, M. A.; Rajagurubandara, E.; Wijesinghe, P.; Bhagwat, A. S., Determinants of sequence-specificity within human AID and APOBEC3G. *DNA repair* **2010,** *9* (5), 579-587.
42. Cen, S.; Guo, F.; Niu, M.; Saadatmand, J.; Deflassieux, J.; Kleiman, L., The interaction between HIV-1 Gag and APOBEC3G. *Journal of Biological Chemistry* **2004,** *279* (32), 33177-33184.
43. Cen, S.; Peng, Z.-G.; Li, X.-Y.; Li, Z.-R.; Ma, J.; Wang, Y.-M.; Fan, B.; You, X.-F.; Wang, Y.-P.; Liu, F., Small molecular compounds inhibit HIV-1 replication through specifically stabilizing APOBEC3G. *Journal of Biological Chemistry* **2010,** *285* (22), 16546-16552.
44. Chabner, B. A.; Roberts, T. G., Chemotherapy and the war on cancer. *Nature Reviews Cancer* **2005,** *5* (1), 65-72.
45. Chabot, G. G.; Bouchard, J.; Momparler, R. L., Kinetics of deamination of 5-aza-2'-deoxycytidine and cytosine arabinoside by human liver cytidine deaminase and its inhibition by 3-deazauridine, thymidine or uracil arabinoside. *Biochemical pharmacology* **1983,** *32* (7), 1327-1328.
46. Chan, K.; Resnick, M. A.; Gordenin, D. A., The choice of nucleotide inserted opposite abasic sites formed within chromosomal DNA reveals the polymerase activities participating in translesion DNA synthesis. *DNA repair* **2013,** *12* (11), 878-889.

47. Chan, K.; Roberts, S. A.; Klimczak, L. J.; Sterling, J. F.; Saini, N.; Malc, E. P.; Kim, J.; Kwiatkowski, D. J.; Fargo, D. C.; Mieczkowski, P. A., An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nature Genetics* **2015,** *47* (9), 1067-1072.
48. Chelico, L.; Pham, P.; Calabrese, P.; Goodman, M. F., APOBEC3G DNA deaminase acts processively 3′→ 5′ on single-stranded DNA. *Nature Structural and Molecular Biology* **2006,** *13* (5), 392-399.
49. Chelico, L.; Pham, P.; Goodman, M. F., Mechanisms of APOBEC3G-catalyzed processive deamination of deoxycytidine on single-stranded DNA. *Nature Structural and Molecular Biology* **2009,** *16* (5), 454-455.
50. Chelico, L.; Sacho, E. J.; Erie, D. A.; Goodman, M. F., A model for oligomeric regulation of APOBEC3G cytosine deaminase-dependent restriction of HIV. *Journal of Biological Chemistry* **2008,** *283* (20), 13780-13791.
51. Chen, H.; Lilley, C. E.; Yu, Q.; Lee, D. V.; Chou, J.; Narvaiza, I.; Landau, N. R.; Weitzman, M. D., APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Current Biology* **2006,** *16* (5), 480-485.
52. Chen, J.-M.; Férec, C.; Cooper, D. N., LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease: mutation detection bias and multiple mechanisms of target gene disruption. *Journal of Biomedicine and Biotechnology* **2006,** *2006*, 1-9.
53. Chen, K.-M.; Harjes, E.; Gross, P. J.; Fahmy, A.; Lu, Y.; Shindo, K.; Harris, R. S.; Matsuo, H., Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* **2008,** *452* (7183), 116-119.
54. Chiu, Y.-L.; Greene, W. C., The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annual Review of Immunology* **2008,** *26*, 317-353.
55. Chiu, Y.-L.; Soros, V. B.; Kreisberg, J. F.; Stopak, K.; Yonemoto, W.; Greene, W. C., Cellular APOBEC3G restricts HIV-1 infection in resting CD4+ T cells. *Nature* **2005,** *435* (7038), 108-114.
56. Chung, S. J.; Fromme, J. C.; Verdine, G. L., Structure of human cytidine deaminase bound to a potent inhibitor. *Journal of medicinal chemistry* **2005,** *48* (3), 658-660.
57. Ciccolini, J.; Dahan, L.; André, N.; Evrard, A.; Duluc, M.; Blesius, A.; Yang, C.; Giacometti, S.; Brunet, C.; Raynal, C., Cytidine deaminase residual activity in serum is a predictive marker of early severe toxicities in adults after gemcitabine-based chemotherapies. *Journal of clinical oncology* **2010,** *28* (1), 160-165.
58. Ciccolini, J.; Mercier, C.; Dahan, L.; André, N., Integrating pharmacogenetics into gemcitabine dosing—time for a change? *Nature reviews Clinical oncology* **2011,** *8* (7), 439-444.
59. Cimmperman, P.; Matulis, D., Protein thermal denaturation measurements via a fluorescent dye. In *Biophysical approaches determining ligand binding to biomolecular targets: detection, measurement and modelling*, Royal Society of Chemistry Biomolecular Sciences: 2011; Vol. 22, pp 247-274.
60. Conticello, S. G., The AID/APOBEC family of nucleic acid mutators. *Genome biology* **2008,** *9* (6), 229.
61. Conticello, S. G.; Harris, R. S.; Neuberger, M. S., The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. *Current Biology* **2003,** *13* (22), 2009-2013.

62. Conticello, S. G.; Langlois, M.-A.; Neuberger, M. S., Insights into DNA deaminases. *Nature Structural and Molecular Biology* **2007,** *14* (1), 7-9.
63. Conticello, S. G.; Thomas, C. J.; Petersen-Mahrt, S. K.; Neuberger, M. S., Evolution of the AID/APOBEC family of polynucleotide (deoxy) cytidine deaminases. *Molecular biology and evolution* **2005,** *22* (2), 367-377.
64. Cordaux, R.; Batzer, M. A., The impact of retrotransposons on human genome evolution. *Nature Reviews Genetics* **2009,** *10* (10), 691-703.
65. Cortez, L. M.; Brown, A. L.; Dennis, M. A.; Collins, C. D.; Brown, A. J.; Mitchell, D.; Mertz, T. M.; Roberts, S. A., APOBEC3A is a prominent cytidine deaminase in breast cancer. *PLoS Genetics* **2019,** *15* (12).
66. Deloch, L.; Derer, A.; Hartmann, J.; Frey, B.; Fietkau, R.; Gaipl, U. S., Modern radiotherapy concepts and the impact of radiation on immune activation. *Frontiers in oncology* **2016,** *6*, 141.
67. Demorest, Z. L.; Li, M.; Harris, R. S., Phosphorylation directly regulates the intrinsic DNA cytidine deaminase activity of activation-induced deaminase and APOBEC3G protein. *Journal of Biological Chemistry* **2011,** *286* (30), 26568-26575.
68. Ding, Q.; Chang, C.-J.; Xie, X.; Xia, W.; Yang, J.-Y.; Wang, S.-C.; Wang, Y.; Xia, J.; Chen, L.; Cai, C., APOBEC3G promotes liver metastasis in an orthotopic mouse model of colorectal cancer and predicts human hepatic metastasis. *Journal of Clinical Investigation* **2011,** *121* (11).
69. Doehle, B. P.; Schäfer, A.; Cullen, B. R., Human APOBEC3B is a potent inhibitor of HIV-1 infectivity and is resistant to HIV-1 Vif. *Virology* **2005,** *339* (2), 281-288.
70. Dutko, J. A.; Schäfer, A.; Kenny, A. E.; Cullen, B. R.; Curcio, M. J., Inhibition of a yeast LTR retrotransposon by human APOBEC3 cytidine deaminases. *Current Biology* **2005,** *15* (7), 661-666.
71. Esnault, C.; Millet, J.; Schwartz, O.; Heidmann, T., Dual inhibitory effects of APOBEC family proteins on retrotransposition of mammalian endogenous retroviruses. *Nucleic acids research* **2006,** *34* (5), 1522-1531.
72. Faguet, G. B., A brief history of cancer: age-old milestones underlying our current knowledge database. *International journal of cancer* **2015,** *136* (9), 2022-2036.
73. Fossat, N.; Tourle, K.; Radziewic, T.; Barratt, K.; Liebhold, D.; Studdert, J. B.; Power, M.; Jones, V.; Loebel, D. A.; Tam, P. P., C to U RNA editing mediated by APOBEC1 requires RNA-binding protein RBM47. *EMBO reports* **2014,** *15* (8), 903-910.
74. FoXS, F., MultiFoXS: single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles Schneidman-Duhovny, Dina; Hammel, Michal; Tainer, John A.; Sali, Andrej. *Nucleic acids research* **2016,** *44*, W424-W429.
75. Franke, D.; Petoukhov, M.; Konarev, P.; Panjkovich, A.; Tuukkanen, A.; Mertens, H.; Kikhney, A.; Hajizadeh, N.; Franklin, J.; Jeffries, C., ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of applied crystallography* **2017,** *50* (4), 1212-1225.
76. Freire, E.; Mayorga, O. L.; Straume, M., Isothermal titration calorimetry. *Analytical Chemistry* **1990,** *62* (18), 950A-959A.
77. Frick, L.; Yang, C.; Marquez, V.; Wolfenden, R. V., Binding of pyrimidin-2-one ribonucleoside by cytidine deaminase as the transition-state analog 3, 4-dihydrouridine and contribution of the 4-hydroxyl group to its binding affinity. *Biochemistry* **1989,** *28* (24), 9423-9430.

78. Fu, Y.; Ito, F.; Zhang, G.; Fernandez, B.; Yang, H.; Chen, X. S., DNA cytosine and methylcytosine deamination by APOBEC3B: enhancing methylcytosine deamination by engineering APOBEC3B. *Biochemical Journal* **2015,** *471* (1), 25-35.
79. Furukawa, A.; Nagata, T.; Matsugami, A.; Habu, Y.; Sugiyama, R.; Hayashi, F.; Kobayashi, N.; Yokoyama, S.; Takaku, H.; Katahira, M., Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G. *The EMBO journal* **2009,** *28* (4), 440-451.
80. Gallois-Montbrun, S.; Holmes, R. K.; Swanson, C. M.; Fernández-Ocaña, M.; Byers, H. L.; Ward, M. A.; Malim, M. H., Comparison of cellular ribonucleoprotein complexes associated with the APOBEC3F and APOBEC3G antiviral proteins. *Journal of virology* **2008,** *82* (11), 5636-5642.
81. Galloway, C.; Kumar, A.; Krucinska, J.; Smith, H., APOBEC-1 complementation factor (ACF) forms RNA-dependent multimers. *Biochemical and Biophysical Research Communications* **2010,** *398* (1), 38-43.
82. Galluzzi, L.; Vitale, I., Driving to cancer on a four-lane expressway. *Trends in Genetics* **2017,** *33* (8), 491-492.
83. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A., ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research* **2003,** *31* (13), 3784-3788.
84. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Wilkins, M. R.; Appel, R. D.; Bairoch, A., Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook*, Springer: 2005; pp 571-607.
85. Gerlinger, M.; Swanton, C., How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *British journal of cancer* **2010,** *103* (8), 1139-1143.
86. Greaves, M.; Maley, C. C., Clonal evolution in cancer. *Nature* **2012,** *481* (7381), 306.
87. Guo, F.; Cen, S.; Niu, M.; Yang, Y.; Gorelick, R. J.; Kleiman, L., The interaction of APOBEC3G with human immunodeficiency virus type 1 nucleocapsid inhibits tRNA3Lys annealing to viral RNA. *Journal of virology* **2007,** *81* (20), 11322-11331.
88. Haché, G.; Liddament, M. T.; Harris, R. S., The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain. *Journal of Biological Chemistry* **2005,** *280* (12), 10920-10924.
89. Hache, G.; Mansky, L. M.; Harris, R. S., Human APOBEC3 proteins, retrovirus restriction, and HIV drug resistance. *AIDS Reviews* **2006,** *8* (3), 148-157.
90. Haché, G.; Shindo, K.; Albin, J. S.; Harris, R. S., Evolution of HIV-1 isolates that use a novel Vif-independent mechanism to resist restriction by human APOBEC3G. *Current Biology* **2008,** *18* (11), 819-824.
91. Hakata, Y.; Landau, N. R., Reversed functional organization of mouse and human APOBEC3 cytidine deaminase domains. *Journal of Biological Chemistry* **2006,** *281* (48), 36624-36631.
92. Harjes, E.; Gross, P. J.; Chen, K.-M.; Lu, Y.; Shindo, K.; Nowarski, R.; Gross, J. D.; Kotler, M.; Harris, R. S.; Matsuo, H., An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. *Journal of Molecular Biology* **2009,** *389* (5), 819-832.
93. Harjes, S.; Jameson, G. B.; Filichev, V. V.; Edwards, P. J. B.; Harjes, E., NMR-based method of small changes reveals how DNA mutator APOBEC3A interacts with its single-stranded DNA substrate. *Nucleic acids research* **2017,** *45* (9), 5602-5613.

94. Harjes, S.; Solomon, W. C.; Li, M.; Chen, K.-M.; Harjes, E.; Harris, R. S.; Matsuo, H., Impact of H216 on the DNA binding and catalytic activities of the HIV restriction factor APOBEC3G. *Journal of virology* **2013,** *87* (12), 7008-7014.
95. Harris R.S.; Petersen-Mahrt S.K.; Neuberger M.S., RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Molecular Cell* **2002,** *10*, 1247–1253.
96. Harris, R. S., Enhancing immunity to HIV through APOBEC. *Nature Biotechnology* **2008,** *26* (10), 1089-1090.
97. Harris, R. S., Cancer mutation signatures, DNA damage mechanisms, and potential clinical implications. *Genome medicine* **2013,** *5* (9), 1-3.
98. Harris, R. S.; Bishop, K. N.; Sheehy, A. M.; Craig, H. M.; Petersen-Mahrt, S. K.; Watt, I. N.; Neuberger, M. S.; Malim, M. H., DNA deamination mediates innate immunity to retroviral infection. *Cell* **2003,** *113* (6), 803-809.
99. Harris, R. S.; Dudley, J. P., APOBECs and virus restriction. *Virology* **2015,** *479*, 131-145.
100. Harris, R. S.; Liddament, M. T., Retroviral restriction by APOBEC proteins. *Nature Reviews Immunology* **2004,** *4* (11), 868-877.
101. Henderson, S.; Chakravarthy, A.; Su, X.; Boshoff, C.; Fenton, T. R., APOBEC-mediated cytosine deamination links PIK3CA helical domain mutations to human papillomavirus-driven tumor development. *Cell reports* **2014,** *7* (6), 1833-1841.
102. Hill, J. M., NMR screening for rapid protein characterization in structural proteomics. In *Structural Proteomics*, Springer: 2008; pp 437-446.
103. Holden, L. G.; Prochnow, C.; Chang, Y. P.; Bransteitter, R.; Chelico, L.; Sen, U.; Stevens, R. C.; Goodman, M. F.; Chen, X. S., Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature* **2008,** *456* (7218), 121-124.
104. Honjo, T.; Muramatsu, M.; Fagarasan, S., AID: how does it aid antibody diversity? *Immunity* **2004,** *20* (6), 659-668.
105. Hopkins, J. B.; Gillilan, R. E.; Skou, S., BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *Journal of applied crystallography* **2017,** *50* (5), 1545-1553.
106. Ikeda, T.; Abd El Galil, K. H.; Tokunaga, K.; Maeda, K.; Sata, T.; Sakaguchi, N.; Heidmann, T.; Koito, A., Intrinsic restriction activity by apolipoprotein B mRNA editing enzyme APOBEC1 against the mobility of autonomous retrotransposons. *Nucleic acids research* **2011,** *39* (13), 5538-5554.
107. Ito, F.; Fu, Y.; Kao, S.-C. A.; Yang, H.; Chen, X. S., Family-wide comparative analysis of cytidine and methylcytidine deamination by eleven human APOBEC proteins. *Journal of Molecular Biology* **2017,** *429* (12), 1787-1799.
108. Ito, F.; Yang, H.; Xiao, X.; Li, S.-X.; Wolfe, A.; Zirkle, B.; Arutiunian, V.; Chen, X. S., Understanding the structure, multimerization, subcellular localization and mC selectivity of a genomic mutator and anti-HIV factor APOBEC3H. *Scientific reports* **2018,** *8* (1), 1-15.
109. Ito, S.; Nagaoka, H.; Shinkura, R.; Begum, N.; Muramatsu, M.; Nakata, M.; Honjo, T., Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proceedings of the National Academy of Sciences USA* **2004,** *101* (7), 1975-1980.
110. Iwatani, Y.; Chan, D. S.; Wang, F.; Maynard, K. S.; Sugiura, W.; Gronenborn, A. M.; Rouzina, I.; Williams, M. C.; Musier-Forsyth, K.; Levin, J. G., Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. *Nucleic acids research* **2007,** *35* (21), 7096-7108.

111. Iwatani, Y.; Takeuchi, H.; Strebel, K.; Levin, J. G., Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect. *Journal of virology* **2006,** *80* (12), 5992-6002.
112. Jafari, R.; Almqvist, H.; Axelsson, H.; Ignatushchenko, M.; Lundbäck, T.; Nordlund, P.; Molina, D. M., The cellular thermal shift assay for evaluating drug target interactions in cells. *Nature Protocols* **2014,** *9* (9), 2100.
113. Jarmuz, A.; Chester, A.; Bayliss, J.; Gisbourne, J.; Dunham, I.; Scott, J.; Navaratnam, N., An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* **2002,** *79* (3), 285-296.
114. Kidd, J. M.; Newman, T. L.; Tuzun, E.; Kaul, R.; Eichler, E. E., Population stratification of a common APOBEC gene deletion polymorphism. *PLoS Genetics* **2007,** *3* (4), e63.
115. Kim, E.-Y.; Bhattacharya, T.; Kunstman, K.; Swantek, P.; Koning, F. A.; Malim, M. H.; Wolinsky, S. M., Human APOBEC3G-mediated editing can promote HIV-1 sequence diversification and accelerate adaptation to selective pressure. *Journal of virology* **2010,** *84* (19), 10402-10405.
116. Kim, E.-Y.; Lorenzo-Redondo, R.; Little, S. J.; Chung, Y.-S.; Phalora, P. K.; Berry, I. M.; Archer, J.; Penugonda, S.; Fischer, W.; Richman, D. D., Human APOBEC3 induced mutation of human immunodeficiency virus type-1 contributes to adaptation and evolution in natural infection. *PLoS pathogens* **2014,** *10* (7), e1004281.
117. Kirby, N.; Cowieson, N.; Hawley, A. M.; Mudie, S. T.; McGillivray, D. J.; Kusel, M.; Samardzic-Boban, V.; Ryan, T. M., Improved radiation dose efficiency in solution SAXS using a sheath flow sample environment. *Acta Crystallographica Section D: Structural Biology* **2016,** *72* (12), 1254-1266.
118. Kitamura, S.; Ode, H.; Nakashima, M.; Imahashi, M.; Naganawa, Y.; Kurosawa, T.; Yokomaku, Y.; Yamane, T.; Watanabe, N.; Suzuki, A., The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. *Nature Structural and Molecular Biology* **2012,** *19* (10), 1005-1010.
119. Knisbacher, B. A.; Gerber, D.; Levanon, E. Y., DNA editing by APOBECs: a genomic preserver and transformer. *Trends in Genetics* **2016,** *32* (1), 16-28.
120. Ko, T.-P.; Lin, J.-J.; Hu, C.-Y.; Hsu, Y.-H.; Wang, A. H.-J.; Liaw, S.-H., Crystal structure of yeast cytosine deaminase insights into enzyme mechanism and evolution. *Journal of Biological Chemistry* **2003,** *278* (21), 19111-19117.
121. Köck, J.; Blum, H. E., Hypermutation of hepatitis B virus genomes by APOBEC3G, APOBEC3C and APOBEC3H. *Journal of General Virology* **2008,** *89* (5), 1184-1191.
122. Kohli, R. M.; Abrams, S. R.; Gajula, K. S.; Maul, R. W.; Gearhart, P. J.; Stivers, J. T., A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase. *Journal of Biological Chemistry* **2009,** *284* (34), 22898-22904.
123. Kohli, R. M.; Maul, R. W.; Guminski, A. F.; McClure, R. L.; Gajula, K. S.; Saribasak, H.; McMahon, M. A.; Siliciano, R. F.; Gearhart, P. J.; Stivers, J. T., Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification. *Journal of Biological Chemistry* **2010,** *285* (52), 40956-40964.
124. Koito, A.; Ikeda, T., Intrinsic immunity against retrotransposons by APOBEC cytidine deaminases. *Frontiers in microbiology* **2013,** *4*, 28.
125. Koning, F. A.; Newman, E. N.; Kim, E.-Y.; Kunstman, K. J.; Wolinsky, S. M.; Malim, M. H., Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. *Journal of virology* **2009,** *83* (18), 9474-9485.

126. Konkel, M. K.; Batzer, M. A. In *A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome*, Seminars in Cancer Biology, Elsevier: 2010; pp 211-221.
127. Koshland Jr, D. E., The application and usefulness of the ratio kcat/KM. *Bioorganic chemistry* **2002,** *30* (3), 211-213.
128. Kouno, T.; Luengas, E. M.; Shigematsu, M.; Shandilya, S. M.; Zhang, J.; Chen, L.; Hara, M.; Schiffer, C. A.; Harris, R. S.; Matsuo, H., Structure of the Vif-binding domain of the antiviral enzyme APOBEC3G. *Nature Structural and Molecular Biology* **2015,** *22* (6), 485.
129. Kouno, T.; Silvas, T. V.; Hilbert, B. J.; Shandilya, S. M.; Bohn, M. F.; Kelch, B. A.; Royer, W. E.; Somasundaran, M.; Yilmaz, N. K.; Matsuo, H., Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity. *Nature communications* **2017,** *8* (1), 1-8.
130. Krause, D. S.; Van Etten, R. A., Tyrosine kinases as targets for cancer therapy. *New England Journal of Medicine* **2005,** *353* (2), 172-187.
131. Krissinel, E., Crystal contacts as nature's docking solutions. *Journal of computational chemistry* **2010,** *31* (1), 133-143.
132. Krissinel, E.; Henrick, K., Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute. *Journal of Molecular Biology* **2007,** *372*, 774-797.
133. Krokan, H. E.; Sætrom, P.; Aas, P. A.; Pettersen, H. S.; Kavli, B.; Slupphaug, G., Error-free versus mutagenic processing of genomic uracil - Relevance to cancer. *DNA repair* **2014,** *19* (1), 38-47.
134. Krzysiak, T. C.; Jung, J.; Thompson, J.; Baker, D.; Gronenborn, A. M., APOBEC2 is a monomer in solution: implications for APOBEC3G models. *Biochemistry* **2012,** *51* (9), 2008-2017.
135. Kvach, M. V.; Barzak, F. M.; Harjes, S.; Schares, H. A.; Jameson, G. B.; Ayoub, A. M.; Moorthy, R.; Aihara, H.; Harris, R. S.; Filichev, V., Inhibiting APOBEC3 activity with single-stranded DNA containing 2'-deoxyzebularine analogs. *Biochemistry* **2019,** *58* (5), 391- 400.
136. Kvach, M. V.; Barzak, F. M.; Harjes, S.; Schares, H. A.; Kurup, H. M.; Jones, K. F.; Sutton, L.; Donahue, J.; D'Aquila, R. T.; Jameson, G. B.; Harki, D. A.; Krause, K. L.; Harjes, E.; Filichev, V., Differential inhibition of APOBEC3 DNA-mutator isozymes by fluoro- and non-fluoro-substituted 2'-deoxyzebularine embedded in single-stranded DNA. *ChemBioChem* **2020,** *21* (7), 1028-1035.
137. Lackey, L.; Demorest, Z. L.; Land, A. M.; Hultquist, J. F.; Brown, W. L.; Harris, R. S., APOBEC3B and AID have similar nuclear import mechanisms. *Journal of Molecular Biology* **2012,** *419* (5), 301-314.
138. Lackey, L.; Law, E. K.; Brown, W. L.; Harris, R. S., Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA deamination. *Cell Cycle* **2013,** *12* (5), 762-772.
139. Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W., Initial sequencing and analysis of the human genome. *Nature* **2001,** *409*, 860--921.
140. Landry, S.; Narvaiza, I.; Linfesty, D. C.; Weitzman, M. D., APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. *EMBO reports* **2011,** *12* (5), 444-450.
141. Langlois, M.-A.; Beale, R. C.; Conticello, S. G.; Neuberger, M. S., Mutational comparison of the single-domained APOBEC3C and double-domained APOBEC3F/G

anti-retroviral cytidine deaminases provides insight into their DNA target site specificities. *Nucleic acids research* **2005,** *33* (6), 1913-1923.
142. LaRue, R. S.; Andrésdóttir, V.; Blanchard, Y.; Conticello, S. G.; Derse, D.; Emerman, M.; Greene, W. C.; Jónsson, S. R.; Landau, N. R.; Löchelt, M., Guidelines for naming nonprimate APOBEC3 genes and proteins. *Journal of virology* **2009,** *83* (2), 494-497.
143. LaRue, R. S.; Jónsson, S. R.; Silverstein, K. A.; Lajoie, M.; Bertrand, D.; El-Mabrouk, N.; Hötzel, I.; Andrésdóttir, V.; Smith, T. P.; Harris, R. S., The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals. *BMC molecular biology* **2008,** *9* (1), 104.
144. Lau, P. P.; Xiong, W.; Zhu, H.-J.; Chen, S.-H.; Chan, L., Apolipoprotein B mRNA editing is an intranuclear event that occurs posttranscriptionally coincident with splicing and polyadenylation. *Journal of Biological Chemistry* **1991,** *266* (30), 20550-20554.
145. Lau, P. P.; Zhu, H.-J.; Baldini, A.; Charnsangavej, C.; Chan, L., Dimeric structure of a human apolipoprotein B mRNA editing protein and cloning and chromosomal localization of its gene. *Proceedings of the National Academy of Sciences USA* **1994,** *91* (18), 8522-8526.
146. Law, E. K.; Sieuwerts, A. M.; LaPara, K.; Leonard, B.; Starrett, G. J.; Molan, A. M.; Temiz, N. A.; Vogel, R. I.; Meijer-van Gelder, M. E.; Sweep, F. C., The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. *Science advances* **2016,** *2* (10), e1601737.
147. Lawrence, M. S.; Stojanov, P.; Polak, P.; Kryukov, G. V.; Cibulskis, K.; Sivachenko, A.; Carter, S. L.; Stewart, C.; Mermel, C. H.; Roberts, S. A., Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **2013,** *499* (7457), 214-218.
148. Lecossier, D.; Bouchonnet, F.; Clavel, F.; Hance, A. J., Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science* **2003,** *300* (5622), 1112-1112.
149. Leonard, B.; Hart, S. N.; Burns, M. B.; Carpenter, M. A.; Temiz, N. A.; Rathore, A.; Vogel, R. I.; Nikas, J. B.; Law, E. K.; Brown, W. L., APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. *Cancer research* **2013,** *73* (24), 7222-7231.
150. Li, J.; Chen, Y.; Li, M.; Carpenter, M. A.; McDougle, R. M.; Luengas, E. M.; Macdonald, P. J.; Harris, R. S.; Mueller, J. D., APOBEC3 multimerization correlates with HIV-1 packaging and restriction activity in living cells. *Journal of Molecular Biology* **2014,** *426* (6), 1296-1307.
151. Li, M.; Shandilya, S. M.; Carpenter, M. A.; Rathore, A.; Brown, W. L.; Perkins, A. L.; Harki, D. A.; Solberg, J.; Hook, D. J.; Pandey, K. K., First-in-class small molecule inhibitors of the single-strand DNA cytosine deaminase APOBEC3G. *American Chemical Society Chemical Biology* **2012,** *7* (3), 506-517.
152. Li, S.-J.; Hochstrasser, M., A new protease required for cell-cycle progression in yeast. *Nature* **1999,** *398* (6724), 246-251.
153. Liao, W.; Hong, S.-H.; Chan, B. H.-J.; Rudolph, F. B.; Clark, S. C.; Chan, L., APOBEC-2, a cardiac-and skeletal muscle-specific member of the cytidine deaminase supergene family. *Biochemical and Biophysical Research Communications* **1999,** *260* (2), 398-404.
154. Liddament, M. T.; Brown, W. L.; Schumacher, A. J.; Harris, R. S., APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo. *Current Biology* **2004,** *14* (15), 1385-1391.

155. Lindahl, T., Suppression of spontaneous mutagenesis in human cells by DNA base excision–repair. *Mutation Research-Reviews in Mutation Research* **2000,** *462* (2), 129-135.
156. Liu, L.; De, S.; Michor, F., DNA replication timing and higher-order nuclear organization determine single-nucleotide substitution patterns in cancer genomes. *Nature communications* **2013,** *4*, 1502.
157. Liu, M.; Mallinger, A. l.; Tortorici, M.; Newbatt, Y.; Richards, M.; Mirza, A.; van Montfort, R. L.; Burke, R.; Blagg, J.; Kaserer, T., Evaluation of APOBEC3B recognition motifs by NMR reveals preferred substrates. *American Chemical Society Chemical Biology* **2018,** *13* (9), 2427-2432.
158. Losey, H. C.; Ruthenburg, A. J.; Verdine, G. L., Crystal structure of Staphylococcus aureus tRNA adenosine deaminase TadA in complex with RNA. *Nature Structural and Molecular Biology* **2006,** *13* (2), 153-159.
159. Lv, W.; Liu, Z.; Jin, H.; Yu, X.; Zhang, L.; Zhang, L., Three-dimensional structure of HIV-1 VIF constructed by comparative modeling and the function characterization analyzed by molecular dynamics simulation. *Organic and Biomolecular Chemistry* **2007,** *5* (4), 617-626.
160. MacDuff, D. A.; Demorest, Z. L.; Harris, R. S., AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity. *Nucleic acids research* **2009,** *37* (6), 1854-1867.
161. Maiti, A.; Myint, W.; Kanai, T.; Delviks-Frankenberry, K.; Rodriguez, C. S.; Pathak, V. K.; Schiffer, C. A.; Matsuo, H., Crystal structure of the catalytic domain of HIV-1 restriction factor APOBEC3G in complex with ssDNA. *Nature communications* **2018,** *9* (1), 2460.
162. Marin, M.; Rose, K. M.; Kozak, S. L.; Kabat, D., HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nature medicine* **2003,** *9* (11), 1398-1403.
163. Marino, D.; Perković, M.; Hain, A.; Vasudevan, A. A. J.; Hofmann, H.; Hanschmann, K.-M.; Mühlebach, M. D.; Schumann, G. G.; König, R.; Cichutek, K., APOBEC4 enhances the replication of HIV-1. *PLoS ONE* **2016,** *11* (6), e0155422.
164. Marley, J.; Lu, M.; Bracken, C., A method for efficient isotopic labeling of recombinant proteins. *Journal of biomolecular NMR* **2001,** *20* (1), 71-75.
165. Martin, K. L.; Johnson, M.; Richard, T., APOBEC3G complexes decrease human immunodeficiency virus type 1 production. *Journal of virology* **2011,** *85* (18), 9314-9326.
166. Marx, A.; Galilee, M.; Alian, A., Zinc enhancement of cytidine deaminase activity highlights a potential allosteric role of loop-3 in regulating APOBEC3 enzymes. *Scientific reports* **2015,** *5*, 18191.
167. McDougall, W. M.; Okany, C.; Smith, H. C., Deaminase activity on single-stranded DNA (ssDNA) occurs in vitro when APOBEC3G cytidine deaminase forms homotetramers and higher-order complexes. *Journal of Biological Chemistry* **2011,** *286* (35), 30655-30661.
168. Mertens, H. D.; Svergun, D. I., Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *Journal of structural biology* **2010,** *172* (1), 128-141.
169. Metzner, M.; Jäck, H.-M.; Wabl, M., LINE-1 retroelements complexed and inhibited by activation induced cytidine deaminase. *PLoS ONE* **2012,** *7* (11), e49358.
170. Miki, Y.; Katagiri, T.; Kasumi, F.; Yoshimoto, T.; Nakamura, Y., Mutation analysis in the BRCA2 gene in primary breast cancers. *Nature Genetics* **1996,** *13* (2), 245-247.

171. Mikl, M. C.; Watt, I. N.; Lu, M.; Reik, W.; Davies, S. L.; Neuberger, M. S.; Rada, C., Mice deficient in APOBEC2 and APOBEC3. *Molecular and Cellular Biology* **2005,** *25* (16), 7270-7277.
172. Mitra, M.; Hercík, K.; Byeon, I.-J. L.; Ahn, J.; Hill, S.; Hinchee-Rodriguez, K.; Singer, D.; Byeon, C.-H.; Charlton, L. M.; Nam, G., Structural determinants of human APOBEC3A enzymatic and nucleic acid binding properties. *Nucleic acids research* **2014,** *42* (2), 1095-1110.
173. Moulder, J. E.; Seymour, C., Radiation fractionation: the search for isoeffect relationships and mechanisms. *International journal of radiation biology* **2018,** *94* (8), 743-751.
174. Mulder, L. C.; Harari, A.; Simon, V., Cytidine deamination induced HIV-1 drug resistance. *Proceedings of the National Academy of Sciences USA* **2008,** *105* (14), 5501-5506.
175. Muramatsu, M.; Kinoshita, K.; Fagarasan, S.; Yamada, S.; Shinkai, Y.; Honjo, T., Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **2000,** *102* (5), 553-563.
176. Muramatsu, M.; Sankaranand, V.; Anant, S.; Sugai, M.; Kinoshita, K.; Davidson, N. O.; Honjo, T., Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *Journal of Biological Chemistry* **1999,** *274* (26), 18470-18476.
177. Muto, T.; Muramatsu, M.; Taniwaki, M.; Kinoshita, K.; Honjo, T., Isolation, tissue distribution, and chromosomal localization of the human activation-induced cytidine deaminase (AID) gene. *Genomics* **2000,** *68* (1), 85-88.
178. Mylonas, E.; Svergun, D. I., Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *Applied Crystallography* **2007,** *40* (s1), s245-s249.
179. Nabel, C. S.; Jia, H.; Ye, Y.; Shen, L.; Goldschmidt, H. L.; Stivers, J. T.; Zhang, Y.; Kohli, R. M., AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nature chemical biology* **2012,** *8* (9), 751-758.
180. Nabel, C. S.; Lee, J. W.; Wang, L. C.; Kohli, R. M., Nucleic acid determinants for selective deamination of DNA over RNA by activation-induced deaminase. *Proceedings of the National Academy of Sciences USA* **2013,** *110* (35), 14225-14230.
181. Nabel, C. S.; Schutsky, E. K.; Kohli, R. M., Molecular targeting of mutagenic AID and APOBEC deaminases. *Cell Cycle* **2014,** *13* (2), 171-172.
182. Nakamuta, M.; Chang, B. H.-J.; Zsigmond, E.; Kobayashi, K.; Lei, H.; Ishida, B. Y.; Oka, K.; Li, E.; Chan, L., Complete phenotypic characterization of apobec-1 knockout mice with a wild-type genetic background and a human apolipoprotein B transgenic background, and restoration of apolipoprotein B mRNA editing by somatic gene transfer of Apobec-1. *Journal of Biological Chemistry* **1996,** *271* (42), 25981-25988.
183. Nathans, R.; Cao, H.; Sharova, N.; Ali, A.; Sharkey, M.; Stranska, R.; Stevenson, M.; Rana, T. M., Small-molecule inhibition of HIV-1 Vif. *Nature Biotechnology* **2008,** *26* (10), 1187.
184. Navaratnam, N.; Bhattacharya, S.; Fujino, T.; Patel, D.; Jarmuz, A. L.; Scott, J., Evolutionary origins of apoB mRNA editing: catalysis by a cytidine deaminase that has acquired a novel RNA-binding motif at its active site. *Cell* **1995,** *81* (2), 187-195.
185. Navaratnam, N.; Fujino, T.; Bayliss, J.; Jarmuz, A.; How, A.; Richardson, N.; Somasekaram, A.; Bhattacharya, S.; Carter, C.; Scott, J., Escherichia coli cytidine deaminase provides a molecular model for ApoB RNA editing and a mechanism for RNA substrate recognition. *Journal of Molecular Biology* **1998,** *275* (4), 695-714.

186. Navaratnam, N.; Morrison, J. R.; Bhattacharya, S.; Patel, D.; Funahashi, T.; Giannoni, F.; Teng, B.; Davidson, N.; Scott, J., The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. *Journal of Biological Chemistry* **1993,** *268* (28), 20709-20712.
187. Navarro, F.; Bollman, B.; Chen, H.; König, R.; Yu, Q.; Chiles, K.; Landau, N. R., Complementary function of the two catalytic domains of APOBEC3G. *Virology* **2005,** *333* (2), 374-386.
188. Neuberger, M. S.; Harris, R. S.; Di Noia, J.; Petersen-Mahrt, S. K., Immunity through DNA deamination. *Trends in biochemical sciences* **2003,** *28* (6), 305-312.
189. Newman, E. N.; Holmes, R. K.; Craig, H. M.; Klein, K. C.; Lingappa, J. R.; Malim, M. H.; Sheehy, A. M., Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Current Biology* **2005,** *15* (2), 166-170.
190. Nik-Zainal, S.; Wedge, D. C.; Alexandrov, L. B.; Petljak, M.; Butler, A. P.; Bolli, N.; Davies, H. R.; Knappskog, S.; Martin, S.; Papaemmanuil, E., Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nature Genetics* **2014,** *46* (5), 487.
191. Ohba, K.; Ichiyama, K.; Yajima, M.; Gemma, N.; Nikaido, M.; Wu, Q.; Chong, P.; Mori, S.; Yamamoto, R.; Wong, J. E. L., In vivo and in vitro studies suggest a possible involvement of HPV infection in the early stage of breast carcinogenesis via APOBEC3B induction. *PLoS ONE* **2014,** *9* (5), e97787.
192. Okazaki, I. m.; Kotani, A.; Honjo, T., Role of AID in tumorigenesis. *Advances in immunology* **2007,** *94*, 245-273.
193. Okuyama, S.; Marusawa, H.; Matsumoto, T.; Ueda, Y.; Matsumoto, Y.; Endo, Y.; Takai, A.; Chiba, T., Excessive activity of apolipoprotein B mRNA editing enzyme catalytic polypeptide 2 (APOBEC2) contributes to liver and lung tumorigenesis. *International journal of cancer* **2012,** *130* (6), 1294-1301.
194. Olson, M. E.; Abate-Pella, D.; Perkins, A. L.; Li, M.; Carpenter, M. A.; Rathore, A.; Harris, R. S.; Harki, D. A., Oxidative reactivities of 2-furylquinolines: ubiquitous scaffolds in common high-throughput screening libraries. *Journal of medicinal chemistry* **2015,** *58* (18), 7419-7430.
195. Olson, M. E.; Harris, R. S.; Harki, D. A., APOBEC enzymes as targets for virus and cancer therapy. *Cell chemical biology* **2017,** *25* (1), 36-49.
196. Olson, M. E.; Li, M.; Harris, R. S.; Harki, D. A., Small-molecule APOBEC3G DNA cytosine deaminase inhibitors based on a 4-Amino-1, 2, 4-triazole-3-thiol scaffold. *ChemMedChem* **2013,** *8* (1), 112-117.
197. Owczarzy, R.; Tataurov, A. V.; Wu, Y.; Manthey, J. A.; McQuisten, K. A.; Almabrazi, H. G.; Pedersen, K. F.; Lin, Y.; Garretson, J.; McEntaggart, N. O., IDT SciTools: a suite for analysis and design of nucleic acid oligomers. *Nucleic acids research* **2008,** *36* (suppl_2), W163-W169.
198. Pace, J. K.; Feschotte, C., The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome research* **2007,** *17* (4), 422-432.
199. Papavasiliou, F. N.; Schatz, D. G., Somatic hypermutation of immunoglobulin genes: merging mechanisms for genetic diversity. *Cell* **2002,** *109* (2), S35-S44.
200. Pery, E.; Sheehy, A.; Nebane, N. M.; Brazier, A. J.; Misra, V.; Rajendran, K. S.; Buhrlage, S. J.; Mankowski, M. K.; Rasmussen, L.; White, E. L., Identification of a novel HIV-1 inhibitor targeting Vif-dependent degradation of human APOBEC3G protein. *Journal of Biological Chemistry* **2015,** *290* (16), 10504-10517.

201. Petersen-Mahrt, S. K.; Harris, R. S.; Neuberger, M. S., AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. *Nature* **2002,** *418* (6893), 99-104.
202. Petersen-Mahrt, S. K.; Neuberger, M. S., In vitro deamination of cytosine to uracil in single-stranded DNA by apolipoprotein B editing complex catalytic subunit 1 (APOBEC1). *Journal of Biological Chemistry* **2003,** *278* (22), 19583-19586.
203. Petit, V.; Guétard, D.; Renard, M.; Keriel, A.; Sitbon, M.; Wain-Hobson, S.; Vartanian, J.-P., Murine APOBEC1 Is a powerful mutator of retroviral and cellular RNA in vitro and in vivo. *Journal of Molecular Biology* **2009,** *385* (1), 65-78.
204. Powell, L. M.; Wallis, S. C.; Pease, R. J.; Edwards, Y. H.; Knott, T. J.; Scott, J., A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell* **1987,** *50* (6), 831-840.
205. Prochnow, C.; Bransteitter, R.; Klein, M. G.; Goodman, M. F.; Chen, X. S., The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature* **2007,** *445* (7126), 447-451.
206. Prohaska, K. M.; Bennett, R. P.; Salter, J. D.; Smith, H. C., The multifaceted roles of RNA binding in APOBEC cytidine deaminase functions. *Wiley Interdisciplinary Reviews: RNA* **2014,** *5* (4), 493-508.
207. Rathore, A.; Carpenter, M. A.; Demir, Ö.; Ikeda, T.; Li, M.; Shaban, N. M.; Law, E. K.; Anokhin, D.; Brown, W. L.; Amaro, R. E., The local dinucleotide preference of APOBEC3G can be altered from 5′-CC to 5′-TC by a single amino acid substitution. *Journal of Molecular Biology* **2013,** *425* (22), 4442-4454.
208. Rausch, J. W.; Chelico, L.; Goodman, M. F.; Le Grice, S. F., Dissecting APOBEC3G substrate specificity by nucleoside analog interference. *Journal of Biological Chemistry* **2009,** *284* (11), 7047-7058.
209. Rayment, I., Diffraction and Scattering by X-Rays and Neutrons. In *Molecular Biophysics for the Life Sciences*, Springer: 2013; pp 91-112.
210. Refsland, E. W.; Harris, R. S., The APOBEC3 family of retroelement restriction factors. In *Intrinsic Immunity*, Springer: 2013; pp 1-27.
211. Refsland, E. W.; Hultquist, J. F.; Harris, R. S., Endogenous origins of HIV-1 G-to-A hypermutation and restriction in the nonpermissive T cell line CEM2n. *PLoS pathogens* **2012,** *8* (7).
212. Refsland, E. W.; Stenglein, M. D.; Shindo, K.; Albin, J. S.; Brown, W. L.; Harris, R. S., Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. *Nucleic acids research* **2010,** *38* (13), 4274-4284.
213. Roberts, S. A.; Lawrence, M. S.; Klimczak, L. J.; Grimm, S. A.; Fargo, D.; Stojanov, P.; Kiezun, A.; Kryukov, G. V.; Carter, S. L.; Saksena, G., An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nature Genetics* **2013,** *45* (9), 970-976.
214. Rogozin, I. B.; Basu, M. K.; Jordan, I. K.; Pavlov, Y. I.; Koonin, E. V., APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy) cytidine deaminases predicted by computational analysis. *Cell Cycle* **2005,** *4* (9), 1281-1285.
215. Rosenberg, B. R.; Hamilton, C. E.; Mwangi, M. M.; Dewell, S.; Papavasiliou, F. N., Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3′ UTRs. *Nature Structural and Molecular Biology* **2011,** *18* (2), 230.

216. Rösler, C.; Köck, J.; Kann, M.; Malim, M. H.; Blum, H. E.; Baumert, T. F.; von Weizsäcker, F., APOBEC-mediated interference with hepadnavirus production. *Hepatology* **2005,** *42* (2), 301-309.
217. Round, A.; Felisaz, F.; Fodinger, L.; Gobbo, A.; Huet, J.; Villard, C.; Blanchet, C. E.; Pernot, P.; McSweeney, S.; Roessle, M., BioSAXS Sample Changer: a robotic sample changer for rapid and reliable high-throughput X-ray solution scattering experiments. *Acta Crystallographica Section D: Biological Crystallography* **2015,** *71* (1), 67-75.
218. Ryan, T. M.; Trewhella, J.; Murphy, J. M.; Keown, J. R.; Casey, L.; Pearce, F. G.; Goldstone, D. C.; Chen, K.; Luo, Z.; Kobe, B., An optimized SEC-SAXS system enabling high X-ray dose for rapid SAXS assessment with correlated UV measurements for biomolecular structure analysis. *Journal of applied crystallography* **2018,** *51* (1), 97-111.
219. Sadler, H. A.; Stenglein, M. D.; Harris, R. S.; Mansky, L. M., APOBEC3G contributes to HIV-1 variation through sublethal mutagenesis. *Journal of virology* **2010,** *84* (14), 7396-7404.
220. Salter, J. D.; Bennett, R. P.; Smith, H. C., The APOBEC protein family: united by structure, divergent in function. *Trends in biochemical sciences* **2016,** *41* (7), 578-594.
221. Sanville, B.; Dolan, M. A.; Wollenberg, K.; Yan, Y.; Martin, C.; Yeung, M. L.; Strebel, K.; Buckler-White, A.; Kozak, C. A., Adaptive evolution of Mus Apobec3 includes retroviral insertion and positive selection at two clusters of residues flanking the substrate groove. *PLoS pathogens* **2010,** *6* (7).
222. Sato, Y.; Probst, H. C.; Tatsumi, R.; Ikeuchi, Y.; Neuberger, M. S.; Rada, C., Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. *Journal of Biological Chemistry* **2010,** *285* (10), 7111-7118.
223. Sawyer, S. L.; Emerman, M.; Malik, H. S., Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS biology* **2004,** *2* (9).
224. Schäffer, A. A.; Aravind, L.; Madden, T. L.; Shavirin, S.; Spouge, J. L.; Wolf, Y. I.; Koonin, E. V.; Altschul, S. F., Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic acids research* **2001,** *29* (14), 2994-3005.
225. Schägger H., Tricine–SDS-PAGE. *Nature Protocols* **2006,** *1* (1), 16-22.
226. Schneidman-Duhovny, D.; Hammel, M.; Tainer, J. A.; Sali, A., Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophysical Journal* **2013,** *105* (4), 962-974.
227. Schramm, V. L., Enzymatic transition state theory and transition state analogue design. *Journal of Biological Chemistry* **2007,** *282* (39), 28297-28300.
228. Schrodinger, LLC, The PyMOL Molecular Graphics System, Version 2.1.1. 2015.
229. Scott, A. M.; Allison, J. P.; Wolchok, J. D., Monoclonal antibodies in cancer therapy. *Cancer Immunity Archive* **2012,** *12* (1), 14.
230. Senavirathne, G.; Jaszczur, M.; Auerbach, P. A.; Upton, T. G.; Chelico, L.; Goodman, M. F.; Rueda, D., Single-stranded DNA scanning and deamination by APOBEC3G cytidine deaminase at single molecule resolution. *Journal of Biological Chemistry* **2012,** *287* (19), 15826-15835.
231. Shaban, N. M.; Shi, K.; Lauer, K. V.; Carpenter, M. A.; Richards, C. M.; Salamango, D.; Wang, J.; Lopresti, M. W.; Banerjee, S.; Levin-Klein, R., The antiviral and cancer genomic DNA deaminase APOBEC3H is regulated by an RNA-mediated dimerization mechanism. *Molecular Cell* **2018,** *69* (1), 75-86. e9.

232. Shaban, N. M.; Shi, K.; Li, M.; Aihara, H.; Harris, R. S., 1.92 Angstrom zinc-free APOBEC3F catalytic domain crystal structure. *Journal of Molecular Biology* **2016,** *428* (11), 2307-2316.
233. Shandilya, S. M.; Nalam, M. N.; Nalivaika, E. A.; Gross, P. J.; Valesano, J. C.; Shindo, K.; Li, M.; Munson, M.; Royer, W. E.; Harjes, E., Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. *Structure* **2010,** *18* (1), 28-38.
234. Sharma, S.; Patnaik, S. K.; Taggart, R. T.; Baysal, B. E., The double-domain cytidine deaminase APOBEC3G is a cellular site-specific RNA editing enzyme. *Scientific reports* **2016,** *6*, 39100.
235. Sharma, S.; Patnaik, S. K.; Taggart, R. T.; Kannisto, E. D.; Enriquez, S. M.; Gollnick, P.; Baysal, B. E., APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. *Nature communications* **2015,** *6* (1), 1-15.
236. Sharma, S.; Wang, J.; Alqassim, E.; Portwood, S.; Gomez, E. C.; Maguire, O.; Basse, P. H.; Wang, E. S.; Segal, B. H.; Baysal, B. E., Mitochondrial hypoxic stress induces widespread RNA editing by APOBEC3G in natural killer cells. *Genome biology* **2019,** *20* (1), 37.
237. Sheehy, A. M.; Gaddis, N. C.; Choi, J. D.; Malim, M. H., Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **2002,** *418* (6898), 646-650.
238. Sheehy, A. M.; Gaddis, N. C.; Malim, M. H., The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nature medicine* **2003,** *9* (11), 1404-1407.
239. Shi, K.; Carpenter, M. A.; Banerjee, S.; Shaban, N. M.; Kurahashi, K.; Salamango, D. J.; McCann, J. L.; Starrett, G. J.; Duffy, J. V.; Demir, Ö., Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nature Structural and Molecular Biology* **2017,** *24* (2), 131.
240. Shi, K.; Carpenter, M. A.; Kurahashi, K.; Harris, R. S.; Aihara, H., Crystal structure of the DNA deaminase APOBEC3B catalytic domain. *Journal of Biological Chemistry* **2015,** *290* (47), 28120-28130.
241. Shi, K.; Demir, Ö.; Carpenter, M. A.; Wagner, J.; Kurahashi, K.; Harris, R. S.; Amaro, R. E.; Aihara, H., Conformational switch regulates the DNA cytosine deaminase activity of human APOBEC3B. *Scientific reports* **2017,** *7* (1), 1-12.
242. Shindo, K.; Li, M.; Gross, P. J.; Brown, W. L.; Harjes, E.; Lu, Y.; Matsuo, H.; Harris, R. S., A comparison of two single-stranded DNA binding models by mutational analysis of APOBEC3G. *Biology* **2012,** *1* (2), 260-276.
243. Shirakawa, K.; Takaori-Kondo, A.; Yokoyama, M.; Izumi, T.; Matsui, M.; Io, K.; Sato, T.; Sato, H.; Uchiyama, T., Phosphorylation of APOBEC3G by protein kinase A regulates its interaction with HIV-1 Vif. *Nature Structural and Molecular Biology* **2008,** *15* (11), 1184.
244. Sieuwerts, A. M.; Willis, S.; Burns, M. B.; Look, M. P.; Meijer-Van Gelder, M. E.; Schlicker, A.; Heideman, M. R.; Jacobs, H.; Wessels, L.; Leyland-Jones, B., Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive breast cancers. *Hormones and Cancer* **2014,** *5* (6), 405-413.
245. Silvas, T. V.; Hou, S.; Myint, W.; Nalivaika, E.; Somasundaran, M.; Kelch, B. A.; Matsuo, H.; Yilmaz, N. K.; Schiffer, C. A., Substrate sequence selectivity of APOBEC3A implicates intra-DNA interactions. *Scientific reports* **2018,** *8* (1), 7511.
246. Simon, V.; Zennou, V.; Murray, D.; Huang, Y.; Ho, D. D.; Bieniasz, P. D., Natural variation in Vif: differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. *PLoS pathogens* **2005,** *1* (1).

247. Siriwardena, S. U.; Guruge, T. A.; Bhagwat, A. S., Characterization of the catalytic domain of human APOBEC3B and the critical structural role for a conserved methionine. *Journal of Molecular Biology* **2015,** *427* (19), 3042-3055.
248. Sivashanmugam A.; Murray V.; Cui C.; Zhang Y.; Wang J.; Li Q., Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. *Protein Science* **2009,** *18*, 936-948.
249. Smith, H. C., The APOBEC1 paradigm for mammalian cytidine deaminases that edit DNA and RNA. In *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution*, CRC Press 2009; pp 181-202.
250. Smith, H. C., RNA binding to APOBEC deaminases; Not simply a substrate for C to U editing. *RNA biology* **2017,** *14* (9), 1153-1165.
251. Sousa, M. M.; Krokan, H. E.; Slupphaug, G., DNA-uracil and human pathology. *Molecular aspects of medicine* **2007,** *28* (3), 276-306.
252. Stenglein, M. D.; Burns, M. B.; Li, M.; Lengyel, J.; Harris, R. S., APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nature Structural and Molecular Biology* **2010,** *17* (2), 222-229.
253. Stenglein, M. D.; Matsuo, H.; Harris, R. S., Two regions within the amino-terminal half of APOBEC3G cooperate to determine cytoplasmic localization. *Journal of virology* **2008,** *82* (19), 9591-9599.
254. Sudhakar, A., History of cancer, ancient and modern treatment methods. *Journal of Cancer Science and Therapy* **2009,** *1* (2), 1-4.
255. Suspène, R.; Aynaud, M.-M.; Guétard, D.; Henry, M.; Eckhoff, G.; Marchio, A.; Pineau, P.; Dejean, A.; Vartanian, J.-P.; Wain-Hobson, S., Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. *Proceedings of the National Academy of Sciences USA* **2011,** *108* (12), 4858-4863.
256. Suspene, R.; Aynaud, M.-M.; Vartanian, J.-P.; Wain-Hobson, S., Efficient deamination of 5-methylcytidine and 5-substituted cytidine residues in DNA by human APOBEC3A cytidine deaminase. *PLoS ONE* **2013,** *8* (6), e63461.
257. Suspène, R.; Guétard, D.; Henry, M.; Sommer, P.; Wain-Hobson, S.; Vartanian, J.-P., Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo. *Proceedings of the National Academy of Sciences USA* **2005,** *102* (23), 8321-8326.
258. Svarovskaia, E. S.; Xu, H.; Mbisa, J. L.; Barr, R.; Gorelick, R. J.; Ono, A.; Freed, E. O.; Hu, W.-S.; Pathak, V. K., Human apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3G (APOBEC3G) is incorporated into HIV-1 virions through interactions with viral and nonviral RNAs. *Journal of Biological Chemistry* **2004,** *279* (34), 35822-35828.
259. Swanton, C.; McGranahan, N.; Starrett, G. J.; Harris, R. S., APOBEC enzymes: mutagenic fuel for cancer evolution and heterogeneity. *Cancer discovery* **2015,** *5* (7), 704-712.
260. Ta, V.-T.; Nagaoka, H.; Catalan, N.; Durandy, A.; Fischer, A.; Imai, K.; Nonoyama, S.; Tashiro, J.; Ikegawa, M.; Ito, S., AID mutant analyses indicate requirement for class-switch-specific cofactors. *Nature immunology* **2003,** *4* (9), 843.
261. Taylor, B. J.; Nik-Zainal, S.; Wu, Y. L.; Stebbings, L. A.; Raine, K.; Campbell, P. J.; Rada, C.; Stratton, M. R.; Neuberger, M. S., DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *elife* **2013,** *2*, e00534.

262. Teng, B.; Burant, C. F.; Davidson, N. O., Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* **1993,** *260* (5115), 1816-1819.
263. Tsimberidou, A.-M., Targeted therapy in cancer. *Cancer chemotherapy and pharmacology* **2015,** *76* (6), 1113-1132.
264. Tuncbag, N.; Gursoy, A.; Nussinov, R.; Keskin, O., Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM. *Nature Protocols* **2011,** *6* (9), 1341.
265. Turelli, P.; Mangeat, B.; Jost, S.; Vianin, S.; Trono, D., Inhibition of hepatitis B virus replication by APOBEC3G. *Science* **2004,** *303* (5665), 1829-1829.
266. Vartanian, J.-P.; Guétard, D.; Henry, M.; Wain-Hobson, S., Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science* **2008,** *320* (5873), 230-233.
267. Vartanian, J.-P.; Henry, M.; Marchio, A.; Suspène, R.; Aynaud, M.-M.; Guétard, D.; Cervantes-Gonzalez, M.; Battiston, C.; Mazzaferro, V.; Pineau, P., Massive APOBEC3 editing of hepatitis B viral DNA in cirrhosis. *PLoS pathogens* **2010,** *6* (5), e1000928.
268. Venkatesan, S.; Rosenthal, R.; Kanu, N.; McGranahan, N.; Bartek, J.; Quezada, S.; Hare, J.; Harris, R.; Swanton, C., Perspective: APOBEC mutagenesis in drug resistance and immune escape in HIV and cancer evolution. *Annals of Oncology* **2018,** *29* (3), 563-572.
269. Vieira, V. C.; Leonard, B.; White, E. A.; Starrett, G. J.; Temiz, N. A.; Lorenz, L. D.; Lee, D.; Soares, M. A.; Lambert, P. F.; Howley, P. M., Human papillomavirus E6 triggers upregulation of the antiviral and cancer genomic DNA deaminase APOBEC3B. *American Society of Microbiology* **2014,** *5* (6), e02234-14.
270. Vlachostergios, P. J.; Faltas, B. M., Treatment resistance in urothelial carcinoma: an evolutionary perspective. *Nature reviews Clinical oncology* **2018,** *15* (8), 495-509.
271. Vonica, A.; Rosa, A.; Arduini, B. L.; Brivanlou, A. H., APOBEC2, a selective inhibitor of TGFβ signaling, regulates left–right axis specification during early embryogenesis. *Developmental biology* **2011,** *350* (1), 13-23.
272. Wagner, J.; Demir, Ö.; Carpenter, M. A.; Aihara, H.; Harki, D. A.; Harris, R. S.; Amaro, R. E., Determinants of Substrate Specificity in APOBEC3B. *ChemRxiv* **2018**.
273. Wang, M.; Rada, C.; Neuberger, M. S., Altering the spectrum of immunoglobulin V gene somatic hypermutation by modifying the active site of AID. *The Journal of experimental medicine* **2010,** *207* (1), 141-153.
274. Wang, X.; Abudu, A.; Son, S.; Dang, Y.; Venta, P. J.; Zheng, Y.-H., Analysis of human APOBEC3H haplotypes and anti-human immunodeficiency virus type 1 activity. *Journal of virology* **2011,** *85* (7), 3142-3152.
275. Warren, C. J.; Pyeon, D., APOBEC3 in papillomavirus restriction, evolution and cancer progression. *Oncotarget* **2015,** *6* (37), 39385-39386.
276. Wedekind, J. E.; Gillilan, R.; Janda, A.; Krucinska, J.; Salter, J. D.; Bennett, R. P.; Raina, J.; Smith, H. C., Nanostructures of APOBEC3G support a hierarchical assembly model of high molecular mass ribonucleoprotein particles from dimeric subunits. *Journal of Biological Chemistry* **2006,** *281* (50), 38122-38126.
277. Wentworth, D. F.; Wolfenden, R., Interaction of 3, 4, 5, 6-tetrahydrouridine with human liver cytidine deaminase. *Biochemistry* **1975,** *14* (23), 5099-5105.
278. Wiegand, H. L.; Doehle, B. P.; Bogerd, H. P.; Cullen, B. R., A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. *The EMBO journal* **2004,** *23* (12), 2451-2458.

279. Willems, L.; Gillet, N. A., APOBEC3 interference during replication of viral genomes. *Viruses* **2015,** *7* (6), 2999-3018.
280. Wong, A.; Soo, R. A.; Yong, W.-P.; Innocenti, F., Clinical pharmacology and pharmacogenetics of gemcitabine. *Drug metabolism reviews* **2009,** *41* (2), 77-88.
281. Xiang, S.; Short, S. A.; Wolfenden, R.; Carter Jr, C. W., Transition-state selectivity for a single hydroxyl group during catalysis by cytidine deaminase. *Biochemistry* **1995,** *34* (14), 4516-4523.
282. Xiao, X. Structural Studies of APOBEC3 Non-Catalytic Domains. Doctoral dissertation, University of Southern California, 2016.
283. Xiao, X.; Yang, H.; Arutiunian, V.; Fang, Y.; Besse, G.; Morimoto, C.; Zirkle, B.; Chen, X. S., Structural determinants of APOBEC3B non-catalytic domain for molecular assembly and catalytic regulation. *Nucleic acids research* **2017,** *45* (12), 7494-7506.
284. Xu, R.; Zhang, X.; Zhang, W.; Fang, Y.; Zheng, S.; Yu, X. F., Association of human APOBEC3 cytidine deaminases with the generation of hepatitis virus B x antigen mutants and hepatocellular carcinoma. *Hepatology* **2007,** *46* (6), 1810-1820.
285. Yamanaka, S.; Balestra, M. E.; Ferrell, L. D.; Fan, J.; Arnold, K. S.; Taylor, S.; Taylor, J. M.; Innerarity, T. L., Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. *Proceedings of the National Academy of Sciences USA* **1995,** *92* (18), 8483-8487.
286. Yang, H.; Ito, F.; Wolfe, A. D.; Li, S.; Mohammadzadeh, N.; Love, R. P.; Yan, M.; Zirkle, B.; Gaba, A.; Chelico, L., Understanding the structural basis of HIV-1 restriction by the full length double-domain APOBEC3G. *Nature communications* **2020,** *11* (1), 1-11.
287. Yang, Y.; Yang, Y.; Smith, H. C., Multiple protein domains determine the cell type-specific nuclear distribution of the catalytic subunit required for apolipoprotein B mRNA editing. *Proceedings of the National Academy of Sciences USA* **1997,** *94* (24), 13075-13080.
288. Yates, L. R.; Campbell, P. J., Evolution of the cancer genome. *Nature Reviews Genetics* **2012,** *13* (11), 795-806.
289. Yoo, C. B.; Jeong, S.; Egger, G.; Liang, G.; Phiasivongsa, P.; Tang, C.; Redkar, S.; Jones, P. A., Delivery of 5-aza-2′-deoxycytidine to cells using oligodeoxynucleotides. *Cancer research* **2007,** *67* (13), 6400-6408.
290. Yu, Q.; König, R.; Pillai, S.; Chiles, K.; Kearney, M.; Palmer, S.; Richman, D.; Coffin, J. M.; Landau, N. R., Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. *Nature Structural and Molecular Biology* **2004,** *11* (5), 435-442.
291. Zhang, H.; Yang, B.; Pomerantz, R. J.; Zhang, C.; Arunachalam, S. C.; Gao, L., The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* **2003,** *424* (6944), 94-98.
292. Zhang, J.; Webb, D. M., Rapid evolution of primate antiviral enzyme APOBEC3G. *Human molecular genetics* **2004,** *13* (16), 1785-1791.
293. Zhang, W.; Zhang, X.; Tian, C.; Wang, T.; Sarkis, P. T. N.; Fang, Y.; Zheng, S.; Yu, X. F.; Xu, R., Cytidine deaminase APOBEC3B interacts with heterogeneous nuclear ribonucleoprotein K and suppresses hepatitis B virus expression. *Cellular microbiology* **2008,** *10* (1), 112-121.
294. Ziegler, S. J.; Hu, Y.; Devarkar, S. C.; Xiong, Y., APOBEC3A loop 1 Is a determinant for single-stranded DNA binding and deamination. *Biochemistry* **2019,** *58* (37), 3838-3847.

295. Zou, J.; Wang, C.; Ma, X.; Wang, E.; Peng, G., APOBEC3B, a molecular driver of mutagenesis in human cancers. *Cell and bioscience* **2017,** *7* (1), 29.
296. Zrenner, R.; Stitt, M.; Sonnewald, U.; Boldt, R., Pyrimidine and purine biosynthesis and degradation in plants. *The Annual Review of Plant Biology* **2006,** *57*, 805-836.
297. Zuo, T.; Liu, D.; Lv, W.; Wang, X.; Wang, J.; Lv, M.; Huang, W.; Wu, J.; Zhang, H.; Jin, H., Small-molecule inhibition of human immunodeficiency virus type 1 replication by targeting the interaction between Vif and ElonginC. *Journal of virology* **2012,** *86* (10), 5497-5507.

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Fareeda Maged Yahya Barzak
Name/title of Primary Supervisor:	Vyacheslav Filichev
Name of Research Output and full reference:	
Kvach, M. V.; Barzak, F. M.; Harjes, S.; Schares, H. A.; Kurup, H. M.; Jones, K. F.; Sutton, L.; Donahue, J.; D'Aquila, R. T.; Jameson, G. B.; Harki, D. A.; Krause, K. L.; Harjes, E.; Filichev, V., Differential inhibition of APOBEC3 DNA-mutator isozymes by	
In which Chapter is the Manuscript /Published work:	Chapter 4
Please indicate:	
• The percentage of the manuscript/Published Work that was contributed by the candidate:	40%
and	
• Describe the contribution that the candidate has made to the Manuscript/Published Work:	
Joint first author. In this article, I expressed and purified the catalytically active A3BCTD variant enzymes (A3BCTD-AL1 and A3BCTD-DM). I quantitatively screened our modified oligonucleotides (3dadZ-oligo, 3dadU-oligo, dZ-oligo, and	
For manuscripts intended for publication please indicate target journal:	
Candidate's Signature:	Fareeda Barzak Digitally signed by Fareeda Barzak Date: 2020.05.26 16:24:47 +12'00'
Date:	26/05/2020
Primary Supervisor's Signature:	Vyacheslav V. Filichev Digitally signed by Vyacheslav V. Filichev Date: 2020.05.26 17:31:36 +12'00'
Date:	26 May 2020

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Fareeda Maged Yahya Barzak
Name/title of Primary Supervisor:	Vyacheslav Filichev
Name of Research Output and full reference:	
Kvach, M. V.; Barzak, F. M.; Harjes, S.; Schares, H. A.; Jameson, G. B.; Ayoub, A. M.; Moorthy, R.; Aihara, H.; Harris, R. S.; Filichev, V., Inhibiting APOBEC3 activity with single-stranded DNA containing 2'-deoxyzebularine analogs. Biochemistry 2019, 5	
In which Chapter is the Manuscript /Published work:	Chapter 4
Please indicate:	
• The percentage of the manuscript/Published Work that was contributed by the candidate:	40%
and	
• Describe the contribution that the candidate has made to the Manuscript/Published Work:	
Joint first author. In this article, I expressed and purified all the A3BCTD variant enzymes (A3BCTD-QM-ΔL3, A3BCTD-QM-ΔL3-E255A, A3BCTD-QM-ΔL3-AL1swap, and A3BCTD-DM). Applying these enzymes, I	
For manuscripts intended for publication please indicate target journal:	
Candidate's Signature:	Fareeda Barzak Digitally signed by Fareeda Barzak Date: 2020.05.26 16:24:47 +12'00'
Date:	26/05/2020
Primary Supervisor's Signature:	Vyacheslav V. Filichev Digitally signed by Vyacheslav V. Filichev Date: 2020.05.26 17:32:14 +12'00'
Date:	26 May 2020

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Fareeda Maged Yahya Barzak
Name/title of Primary Supervisor:	Vyacheslav Filichev
Name of Research Output and full reference:	
Barzak, F. M.; Harjes, S.; Kvach, M. V.; Kurup, H. M.; Jameson, G. B.; Filichev, V. V.; Harjes, E., Selective inhibition of APOBEC3 enzymes by single-stranded DNAs containing 2deoxyzebularine. Organic & biomolecular chemistry 2019, 17 (43), 9435-9	
In which Chapter is the Manuscript /Published work:	Chapter 5
Please indicate:	
• The percentage of the manuscript/Published Work that was contributed by the candidate:	70%
and	
• Describe the contribution that the candidate has made to the Manuscript/Published Work:	
First author. In this article, I hypothesised and tested whether selective inhibition using the differentially deaminated suboptimal CCC-oligo was feasible. First, I expressed and purified the A3BCTD variant enzymes (A3BCTD-QM-AL3-AL1swap	
For manuscripts intended for publication please indicate target journal:	
Candidate's Signature:	Fareeda Barzak Digitally signed by Fareeda Barzak Date: 2020.05.26 16:24:47 +12'00'
Date:	26/05/2020
Primary Supervisor's Signature:	Vyacheslav V. Filichev Digitally signed by Vyacheslav V. Filichev Date: 2020.05.26 17:30:58 +12'00'
Date:	26 May 2020

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)